R E P O R T   R E S U M E S

ED 018 974   24   EF 001 648

SIZE FACTORS AND NON-DOLLAR COSTS OF SECONDARY SCHOOLS, PHASE 1. FINAL REPORT.

BY- WAKEFIELD, HOWARD E.   AND OTHERS

PUB DATE   APR 68

CONTRACT OEC-3-6-061266-0801

EDRS PRICE   MF-$1.00   HC-$8.88   220P.

DESCRIPTORS- *ACHIEVEMENT, *COMMUNITY CHARACTERISTICS, *PRODUCTIVITY, *SCHOOL DEMOGRAPHY, *SCHOOL SIZE, CONCEPTUAL SCHEMES, ORGANIZATION SIZE (GROUPS), SECONDARY SCHOOLS, UNIVERSITY OF WISCONSIN, IOWA, FLORIDA

ELEMENTS OF EDUCATIONAL PRODUCTIVITY ARE IDENTIFIED WHICH ARE INFLUENCED BY SIZE OF SECONDARY SCHOOLS. PRODUCTIVITY INCLUDES OUTCOMES SUCH AS ACADEMIC PROGRESS AND SOCIAL BEHAVIOR AND EXCLUDES ECONOMIC OUTCOMES SUCH AS GAINS IN INCOME. RESEARCH IN THIS FIRST PHASE WAS DONE WITH THE HIGH SCHOOLS OF IOWA. DATA STORED WITH THE IOWA EDUCATIONAL INFORMATION CENTER AND THE U.S. BUREAU OF THE CENSUS WERE USED. CLASSES OF INDICATORS WERE DEVELOPED AND ANALYZED STATISTICALLY. STEPS WERE THEN TAKEN TO (1) DEFINE SUB-CLASSES OF HOMOGENEOUS INDICATORS, (2) PERFORM PRINCIPAL COMPONENTS ANALYSIS ON EACH SUB-CLASS TO CHECK HOMOGENEITY, REMOVE AMBIGUITY AND COMPUTE COMPONENT SCORES, (3) PERFORM IMAGE ANALYSIS ON THE COMPONENTS WITHIN EACH CLASS OF INDICATORS AND COMPUTE IMAGE FACTOR SCORES, (4) PERFORM REGRESSION ANALYSIS OF IMAGE FACTORS FOR EACH CLASS OF INDICATORS USING PRODUCTIVITY AS THE DEPENDENT VARIABLE, AND (5) INTERPRET SIGNIFICANT INTERACTION REGRESSION COEFFICIENTS. TEN IMAGE FACTORS WERE EXTRACTED, BUT THE SPARSITY OF DATA ON SOCIAL CHARACTERISTICS OF SCHOOL DISTRICTS REDUCED THE VALIDITY OF THE ANALYSIS. A PRINCIPAL CONCLUSION OF THIS PHASE IS THAT DATA DESIGNED AND COLLECTED FOR ADMINISTRATIVE PURPOSES CAN BE AS COSTLY AND TIME CONSUMING TO CONVERT TO RESEARCH USES AS THE SAMPLING OF POPULATIONS AND GATHERING OR ORIGINAL DATA. (BD)

BR-6-1266
P.A. 24

ED018974

FINAL REPORT PA-24
Project No. 6-1266
Contract No. OEC 3-6-061266-0801

SIZE FACTORS AND NON-DOLLAR COSTS OF SECONDARY SCHOOLS, PHASE I

April 1968

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

EF 001648

FINAL REPORT

Project No. 6-1266
Contract No. OEC 3-6-061266-0801

SIZE FACTORS AND NON-DOLLAR COSTS OF SECONDARY SCHOOLS, PHASE I

Howard E. Wakefield
Donald M. Miller
Richard G. Wolfe

University of Wisconsin

Madison, Wisconsin

April 1968

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

PREFACE

This research project grew out of an operational concern of a local public school district. A member of the school board had raised a question about comparative costs of operating two 1500-pupil high schools vs one 3,000 pupil high school. While lending assistance in resolving the operational problem, members of the University faculty began to see a much broader area of concern. Several costs are paid when decisions are made about the size of educational institutions, only one of which is the number of dollars.

Many, many people have been involved in the conduct of this research. Dr. David E. Wiley, University of Chicago, and Dr. David Nasitir, University of California, were consulted in the development of the proposal and periodically during the course of the research. Several other consultants were involved at critical points in the research. Included among them are Dr. Charles M. Bonjean, University of Texas, Professor Randall Sale, University of Wisconsin, and Dr. Thomas J. Johnson, Central Midwestern Regional Educational Laboratory.

Proper recognition is due also to individuals in Iowa, Florida and the U.S. Census Bureau. In Iowa, three persons in particular should be recognized. The first person is the superintendent of public instruction, Mr. Paul Johnston, who authorized the use of Iowa data and assigned staff to assist the research team. In that capacity, Mr. D. J. Gilliland and Dr. Ralph Van Dusseldorp were our contacts in the Department of Public Instruction and in the Iowa Educational Information Center.

In like fashion, Floyd T. Christian, the state superintendent in Florida, authorized our use of their data. Our earlier contacts in Florida were with the late Dr. Robert W. Sims and, following his death, Dr. Everett Yarbrough and Dr. Archie B. Johnston.

The Population Division of the U.S. Census Bureau was helpful in providing us with census data for the states involved. Our first contacts were with Mr. Herman P. Miller and subsequently with Mr. Louis H. Conger, Jr., who continued to work with us throughout the remainder of the study.

The work of a number of persons on the project staff should be recognized. Mrs. Jane D. Armstrong was responsible for analyzing the content of the Iowa and Florida data bank format books. Jacob Feldman assisted in the preparation of a bibliography of organizational size and in the analysis of the 48 contiguous states. John Proctor assisted in the generation of a conceptual framework within which to analyze school size. Mr. Roger Voytecki was responsible for general programming and for the machine indexing of the format books. Brother Francis Wray assisted with the inquiry into social area analysis and with the Iowa mapping problem.

Use of the University of Wisconsin Computing Center was made possible through the University of Wisconsin Research Committee due to support, in part, from the National Science Foundation, United States Government Agencies and the Wisconsin Alumni Research Foundation (WARF).

Professor Stewart D. North was instrumental in the early identification of the research problem and the development of the proposal and is identified as a co-investigator in the project. While a special assignment as director of a job corp educational program at Sparta, Wisconsin prevented him from participating actively in the research, he continued to advise and encourage the research team.

It is apparent from the above that this research project has required the energies and talents of many people. To all of the above and to many others unnamed who in one way or another assisted the project we express our indebtedness and gratitude.

H.E.W.
D.M.M.
R.G.W.

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

LIST OF FIGURES

Figure		Page

CHAPTER I

THE NATURE OF THE PROBLEM AND THE GENERAL RESEARCH OBJECTIVES

This research arose out of the concern of one school district about the relative costs of varying sizes of high schools. The immediate problem was with dollar costs: would it be more economical, in terms of construction and maintenance costs per pupil, to build one large school or two smaller ones. It was apparent almost at once that other costs of size are incurred. These other or "non-dollar" costs are less obvious but no less serious in that they may deplete our human resources in ways economic efficiency cannot overcome. The non-dollar costs concern the quality of life and education in a school and the educational productivity of that school in terms of the general achievement of the students.

On one hand, increased school size means increased population density and organizational complexity. It seems logical, for example, that higher concentrations of pupils within the confined setting of a school building would result in higher energy drains and that the complex organization of a large school would alter the personal sense of worth of individuals and emphasize instead the worth of the organization. The non-dollar costs would be reflected in such immediate ways as reduced productivity in tackling and mastering learning tasks and in such ultimate ways as social or psychological maladjustment. On the other hand, increased size means diversity of educational and social activity. For example, there may be more varied course work offered in a large school and there may be more possibility for meeting new and different people. The non-dollar value would then be reflected in an increase of actual and potential fulfillment.

It seems, however, that the economic consequences of our acts and decisions in designing schools evoke instant interest, but the non-economic consequences are felt to be somewhat innocuous and therefore less disturbing. Thus the study of the non-economic consequences--of their causes and of their effects--has, in general, been neglected. We are left with few ways of being warned of the consequences, with fewer clues to their onset, and with a nagging uncertainty about the level of educational quality which can reasonably be expected with varying high school size.

These considerations might have been generalized to all aspects of the learning environment, but high school size was clearly a decision frequently made and changed with little regard for anything except number of pupils, square or cubic footage of space, and financial resources. The research reported herein was designed, then, to relate size of secondary school to the non-dollar costs of education carried out within the environment established by size. Specifically, the research was directed toward the following four objectives:

1. To evolve a substantive theoretical framework in which empirical investigation of the influence of size on the interrelationships of community, school, and educational output can be carried out;

2. To develop and implement a methodological and multivariate statistical scheme in which the framework may be represented and applied to data;

3. To seek, obtain, and organize the relevant data which are available in machine-readable form, for two states (Iowa and Florida); and

4. To examine the utility of these kinds of data, of this kind of methodology, and of the theoretical framework in arriving at educationally meaningful results.

The ultimate goal or outgrowth of this research would be a set of substantive relationships useful in gauging the influence that the size of an educational plant has on educational productivity: a local school board should through some algorithm be able to estimate the educational consequences of its actions and decisions. This research was intended to test an approach to achieving that goal.

In Section A of this chapter of the report, an initial conceptualization of the relationships between school size and community characteristics, educational processes, and educational productivity is presented. In Section B, the methodological approach to measuring theoretically-defined categories is described. In Section C, the kinds of data available and the problems in using them are outlined.

A. School Size and the Educational Process

The research reported here was not experimental, for it would have been impossible to have performed a true experiment testing the effect of school size: it would have been impossible to have selected school areas randomly, built schools of systematically different sizes, and measured the output productivities. Instead, this research consisted of a survey of the current educational situations in a large number of communities and a comparison of the measurable community, school, and productivity characteristics with the actual school sizes.

The great problem with surveys, as opposed to proper experiments, is in ascertaining which of the derived relationships can reasonably be considered causal. There must be strong theoretical justification for any inference made. Suppose, for example, that mean occupational aspiration level were correlated with school size: this is unenlightening unless there is strong theoretical justification for size influencing occupational choice, e.g., by acquainting students with suitable occupational choices. In the complex data considered here, there are hundreds of such relationships; some are causal and others are not. A theoretical framework is necessary in accessing the educational significance of the results of analysis. It is also essential to the systematic design of the analysis; i.e., there must be theoretical guidelines for deciding what variables should be constructed and how they should be compared. This is especially true when, as in the present research, there is a large range of possible variables and analyses. For example, because there were so many data items, some item composition had to be performed, and a theoretical orientation was necessary in deciding which items should be composited.

The metatheoretical notions for the research--that is, the goals and format used in generating the theory--are presented in this section. The details and formal definitions of the theory as finally evolved are given in Chapter 5. It should also be noted that one reason for performing the auxiliary survey of U.S. educational climate, presented in Chapter 3, was to provide a small example of the metatheoretical techniques which were later used in the main analysis of high school education in Iowa.

The first metatheoretical concept is that of a <u>cluster.</u> Clusters are qualities or processes which are considered to interact with one another in a primary causal way. For example, the process of "stimulation by conversation with peers about current events" may cause the quality "political awareness" to be acquired. On the other hand, a quality such as "color of classroom walls" is not the primary cause of anything important except perhaps through such intervening causes as "state of being". To say that a cluster, a process or quality exists is not to say that it relates to a particular variable or set of variables which one has in mind, for, as noted later, there may be no direct measurement of a cluster. But a metatheoretical rule is that a cluster should be essentially a scale, and that a school should have a large or small amount of it.

There are in the complex content domain associated with the problem of high school size a multitude of primary causes and hence of clusters. It is necessary therefore to organize further. A related group of clusters is said to be a <u>supercluster</u>. Because this research is concerned with assigning causes, the organization of the clusters has been along the lines of the agents and recipients of causes. For example, all educational and social processes that take place in a school, which is an agent of mediation, are considered together in a supercluster. The primary processes are grouped so that cause, in the sense of responsibility, can be assessed.

The superclusters adopted for this research are diagrammed in Figure 1-A-1. The details are given in Chapter 5. As mentioned, the set of processes occurring within a school is considered to form a supercluster, and in the diagram this is labelled "School Mediation". The size of a school is treated as a special separate supercluster in the scheme: it is labelled "School Manipulation", which is intended to suggest that size, along with other possible objects of study, can be manipulated by administrators in devising a school environment. While school size does not directly affect the students of a school, it helps determine the more fundamental processes which do, namely,

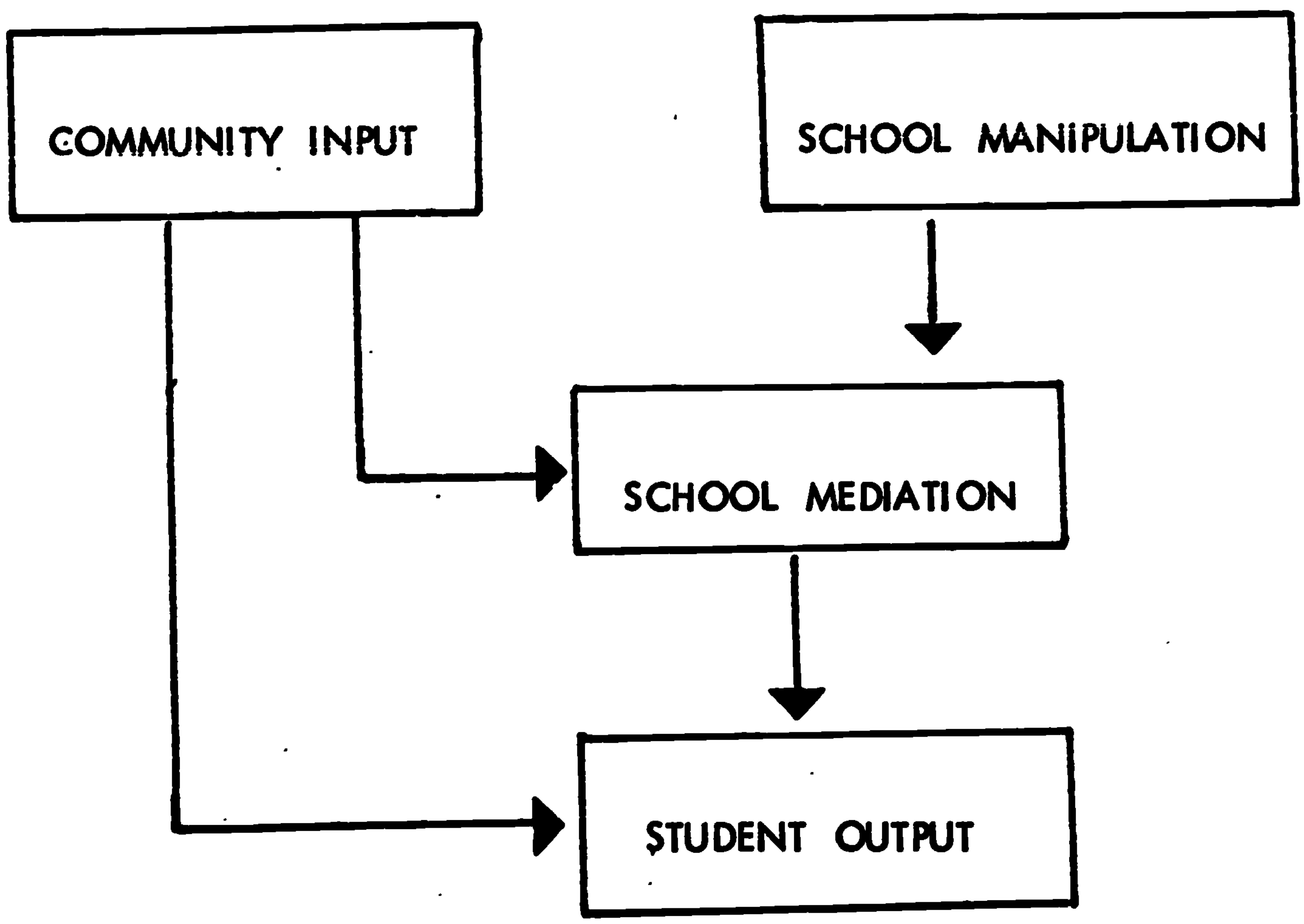


Figure 1-A-1

The Superclusters of the Theory

"School Mediation". The students of a school are represented by the "Community Input" and "Student Output" superclusters; i.e., a school receives a student body with certain input characteristics, determined by the community, and after mediation yields the output student body. This may be conceived as involving a transformation of the values of the same clusters. The additional line connecting "Community Input" and "Student Output" indicates the continuing influence of the community over the growth of a student while he is in school.

In order to establish sufficiently realistic descriptive complexity of the educational situation, it is necessary to define many clusters even within a single supercluster. Therefore, in order to provide adequate parsimony, a further metatheoretical notion is introduced--that of factors. A factor is a pattern of clusters within a supercluster which in actual data is found to be constant across schools; i.e., while clusters represent single theoretical causal qualities and processes, factors provide a reduction in the consideration of them. This includes consideration of their total variation but is simpler to use. For example, if two processes, say "stimulation in advanced mathematics" and "stimulation in advanced physics" are found always to occur together, it is reasonable and does not imply a loss of information to reduce them to a single factor such as "stimulation in advanced science". The differential relationships of a set of clusters can be much more complex, but the general idea is that of empirical reduction.

To summarize, the theory is developed in order to provide justification for assuming causality in the derived relationships and for guiding the analytic procedures. In the metatheory, there are clusters, corresponding to qualities and processes of primary cause, grouped into superclusters, corresponding to agents or recipients of cause. And within each supercluster the array of clusters is reduced to factors, representing empirically-found patterns of cluster variation. In analyzing and interpreting according to the theory, comparisons are made between the factors of one supercluster and the factors of another, while controlling for the effects of factors of still other superclusters.

B. Measuring the Effect of School Size

Given the theoretical framework outlined in the previous section, the measurement problems are threefold: to obtain variables corresponding to the clusters, to combine those variables into factors, and to compare the factors. Because of the restriction to data already available, the problem of determining variables for a cluster could not be solved by ascertaining the most appropriate test in terms of face validity. Instead, indicators of the clusters had to be found. An indicator is a data item which is expected to correlate with a cluster but which is not necessarily equivalent in definition or even causation with the cluster. The values of an indicator are merely required to be associated with the hypothetical values achieved for a cluster. For example, in measuring the cluster "intellectual satisfaction", indicators might be "correlation between aptitude tests and grades" (when there is a low correlation, some students are frustrated in the learning process) and "library usage" (high library usage may mean students enjoy reading). The assignment of indicators to clusters must be justified in terms of the theoretical orientation.

Because an indicator for a cluster is not in general a direct measure of the cluster, it usually contains spurious variance; i.e., the variance of an indicator may be considered in three parts: cluster variance, which corresponds to measurement of the cluster; error variance, which corresponds to imprecise determination of the indicator; and indicator variance, which represents the substantive difference between the actual indicator and the ideal cluster. The amount of indicator variance present in the measurement of an indicator must be minimized, for it may not only lower the precision of the later analysis, just as the error variance does, but it may also substantively bias the later analysis, for to the extent that indicator variance is present, the variable constructed for a cluster does not mean what the cluster does. The approach taken to lower the indicator variance involves determining several indicators for each cluster and compositing them into a single score. The idea is that when several indicators are combined, their common parts are reinforced and their indicator parts are cancelled.

The method of composition used in this research was principal component analysis (Hotelling, 1933). Given a set of items, the principal component is the linear combination of them which has the property that the sum of squares of the correlations of the item with the combination score is maximized, which is to say that the maximum amount of variance in the items is extracted. When the items are relatively homogeneous--when they have high intercorrelations--the amount of variance extracted by the principal component is relatively high. It should be noted that only the unique principal component is used; the analysis is not used here to factor the clusters.

Given a set of principal component scores which represent the clusters, the next problem is to combine them into factors within superclusters. For this purpose image analysis (Guttman, 1953) was used. Image analysis is a hybrid of components analysis and pure factor analysis. It involves an implicit transformation of the original variables into a set of image variables--each original variable is replaced by its linear regression prediction from the other variables--and then a components analysis is made of the transformed variables. The transformation provides variables which approximate the common variables of pure factor analysis, but the subsequent components analysis allows exact computation of factor scores. In this research, a modification (Harris, 1962) of Guttman's original formulation of image analysis is used. This modification sharpens the correspondence to pure factor analysis. Because the image factors are to be given interpretation in the later regression, they are rotated to simple structure. The normal varimax orthogonal rotation scheme (Kaiser, 1958) is used.

Two recent studies of school districts have profitably used image analysis. Peterson, Rossmiller, North, and Wakefield (1963) studied the relationships among 131 indicators of the financial status of school districts in Wisconsin. The indicators used included, for example, assessed valuation per capita, mean family income, pupils per teacher, and the percent of the labor force in various categories. Twenty general factors were identified: twelve of these characterized financial ability of a district while eight referred to financial need. Miller, Conry, Wiley, and Wolfe (1967) studied the interrelationships among characteristics of elementary school districts in Wisconsin. These characteristics included nineteen demographic indicators such as number of elementary students enrolled, number of one-room schools in the district, and valuation per school. Twelve other characteristics were indices of teacher qualities, such as the mean and variance of teacher salaries, local and total experience, and highest credential held.

The relationships between the superclusters--input, mediation, output, and size--are examined with multiple regression analysis (see, e.g., Goldberger, 1963) on the factors across the superclusters. Two special regression techniques are used. First, the specific contribution of each independent factor is considered separately. This is possible because the factors within a supercluster are orthogonal; i.e., one may determine how much variance each factor accounts for without worrying about correlation of effects. Second, a regression technique not usually found in educational research is appropriate here; this is sometimes called "two-stage least-squares" and involves computing regression equations between two sets of variables after having controlled or partialled out the effect of a third set. For example, in regressing School Mediation on Student Output, the Community Input factors may be partialled.

Multiple regression analysis has been employed by Thomas (1962) in consideration of variables selected from Project TALENT data. Thomas investigated the relationships among sixteen indicators of the socioeconomic class of communities and sixteen indicators describing qualities of teacher staff. He found, for example, that beginning salaries and teaching experience were significantly related to student achievement and that the age of a school is positively related to student test scores.

The general flow of the analysis and measurement is diagrammed in Figure 1-B-1. The ultimate regression study is neither a single hypothesis test nor a simple slope determination: the analysis throughout is multivariate, and the effect of school size is considered to involve complex interaction of input and mediational factors with output factors. In that respect, the actual computation presented in this report in Chapter 5 represents only the beginning of analysis of the present data.

The theoretical framework--cluster, supercluster, factor--and the methodological framework--indicator, composite, factor--for this research are rather original. Ordinarily one finds studies in the literature of schools and social area analysis in which raw variables, more or less carefully picked, are regressed. Less often, the variables are factored and the factors are compared. Several objections to such approaches led to the development of the present study.

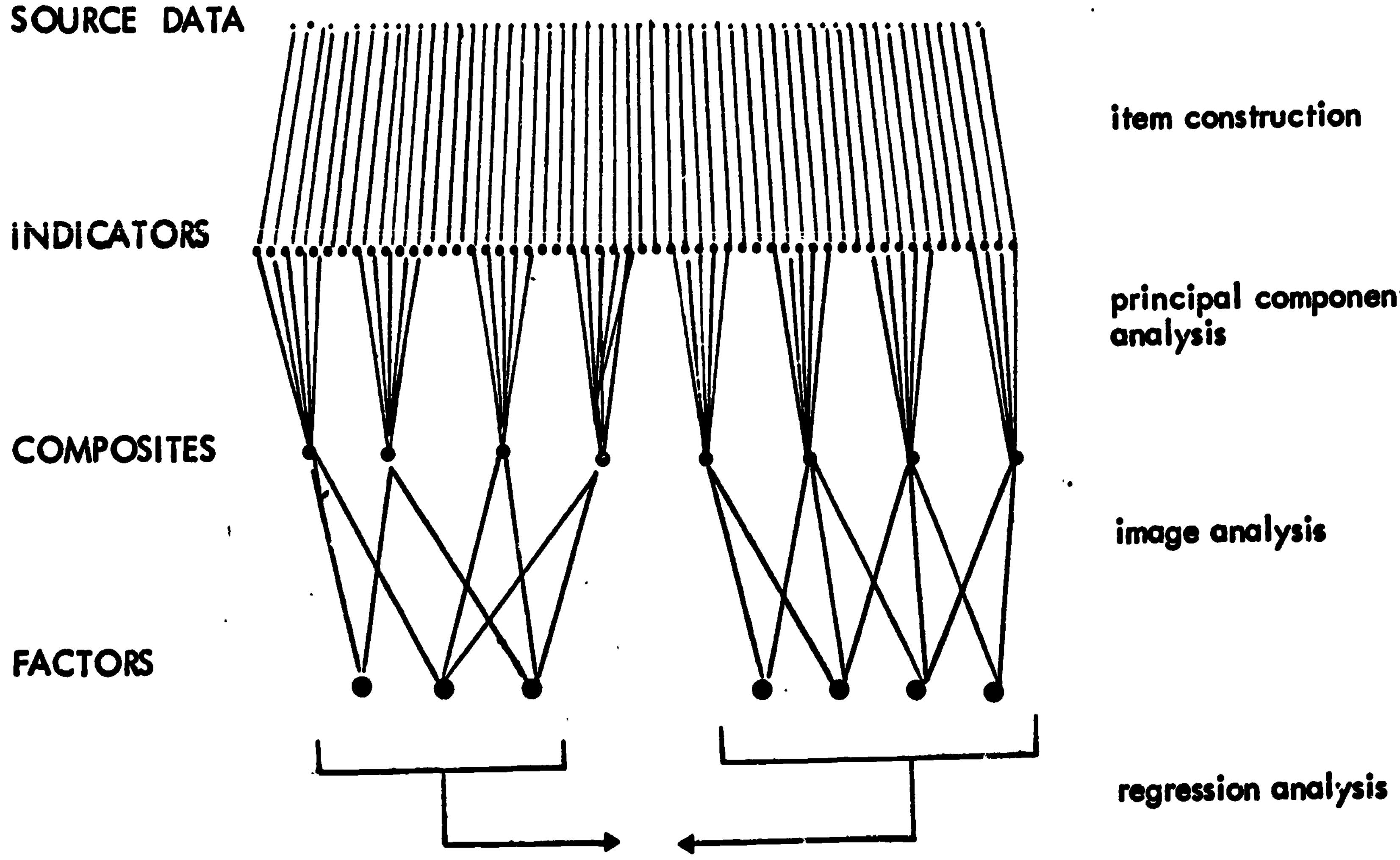


Figure 1-B-1

Data Reduction

The importance and meaningfulness of almost any particular raw demographic or educational data item can be questioned. This is not a matter of error; it is a matter of face validity. For example, the wealth of a community is not really the same as the median income of its inhabitants; saying that a community is wealthy means both more and less than saying the inhabitants have a high median income; saying a school has good teachers is not the same as saying the average credential of the teachers is high. In analyzing demographic and school data, it is essential to remember that the raw items are indicators, not essential causal variables. Thus, in the present research, compositing of indicators was performed to achieve a higher degree of meaningfulness.

Factorization is no help in bringing meaningfulness to a set of variables. A factor is computed when several items are associated in the values that they achieve over the input sample. There is no assurance that such association is causal or meaningful. For example, schools with advanced mathematics courses usually have advanced physics courses, but having an advanced mathematics course does not imply that anyone is going to learn advanced physics. Factorization may even be dangerous in analyzing demographic and school variables, because the resulting factors are often neat and satisfying. This is especially true when a large mass of variables is reduced to a relatively small number of factors. The output from a factor analysis can be no more meaningful than the input, and unless variables are chosen to measure significant theoretical processes and qualities, the outputs are useless in providing an explanation for the causal situation. For example, when two highly correlating variables are included in an analysis they will almost always form a factor, yet they may be related to entirely different theoretical constructs. In this research, the compositing operations were intended to provide meaningful variables for analysis.

C. Obtaining and Manipulating Data

The research was confined to the use of data already available in machine-readable form. It therefore became, in part, a test of whether such data are adequate for obtaining substantive results expected of exacting research inquiries. The two major sources of data were: the U.S. Census Bureau for socio-economic data, and state departments of education for educational data. In both instances, the data were originally gathered and stored for other purposes. Of the five kinds of data sought, three were available in the Iowa and Florida educational data banks--namely, school, school district, and teacher data. The Iowa data bank also had complete data on students, while Florida had data on only a sample of the students. The Census Bureau data bank had data on all areas in both Iowa and Florida.

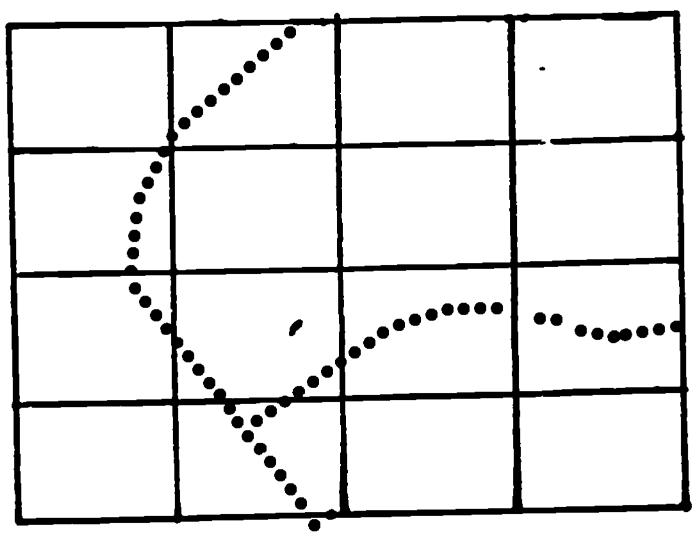
A preliminary problem in using the data banks was to document the data items available for all entities and to determine the format of the data. In the preliminary stages of the research, a complete content analysis of the data bank manuals was made and a computer-made index was prepared. This process took a surprisingly large amount of time and effort. A major concern in beginning the analysis of the data was with the alignment of different data units. For example, students are units within schools, and school attendance areas are units which may differ from census tracts. Procedures had to be worked out for dealing with such hierarchical and disconsonant units.

The problem of hierarchical units is a very difficult one statistically. No adequate statistical technique was found which provided for comparing data at different levels in a hierarchy. How can teacher quality be compared adequately with student performance? The problem is really one of relating the higher unit's variables to the distribution of the lower unit's variables, e.g., teacher quality is related to the distribution of student performance. The simplest way around this difficulty is to compare some computed characteristic of the distributions at the lower level with the variables at the higher level, e.g., comparing teacher quality with mean performance. This and an associated technique were used in this study. The associated technique consisted of taking an additional characteristic--namely, the variance of the distribution at the lower level--and comparing it.

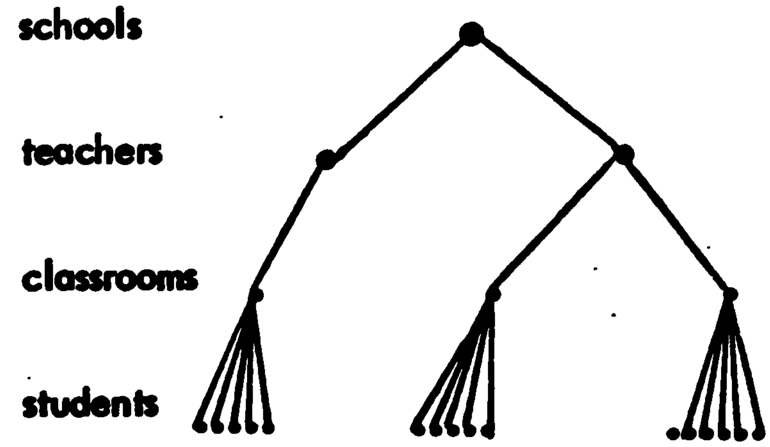
The problem of disconsonance is very difficult practically. This problem arose in Iowa when the school areas were to be compared with the census units. There is essentially no correspondence. For the purpose of this research all data had to be made. First, methods for approximating the correspondence with statistical adjustment were considered. Finally, the exhausting job of comparing maps and assigning census units to school districts proportionally by area had to be performed.

The problems of making correspondence between hierarchical and disconsonant units is diagrammed in Figure 1-C-1. The substantive validity of using such data banks is discussed in Chapter 5. It is appropriate to note here their use for purposes other than originally intended creates difficulties not easily foreseen.

EXAMPLE OF
DISCONSONANT UNITS



EXAMPLE OF
HIERARCHICAL UNITS



. . . . school district lines

——— census division lines

Figure 1-C-1
Hierarchical and Disconsonant Units

CHAPTER 2

PERSPECTIVES ON THE RESEARCH PROVIDED BY PREVIOUS INVESTIGATORS

In order to provide a perspective for the present research on high school size as related to educational productivity and to identify substantive constructs and hypotheses relevant to the topic, a survey of pertinent literature was undertaken. The findings that previous investigators have found to be related to organizational size are reported in Section A of this chapter, along with a tabulation of the variables. In Section B, a resume of research directly related to school size is presented. As was explained in Chapter 1, school size is one link in the causal chain of the educational process. Related to this is the fact that while community characteristics do exert an important effect upon students, the effect of school size is differential across types of communities. Therefore, in Section C, a survey of the literature on social area analysis and a tabulation of the variables which previous researchers have found relevant to this topic are presented. Finally, in Section D, there appears an abbreviated transcript of a colloquium on the Determination of School-Community Characteristics which was held in January, 1967. As the transcript indicates, the meeting was initially designed to focus on the problems of measuring the community characteristics of school areas. However, the discussion proceeded to probe into the problems and relevance of planned measurement as it concerned this specific research undertaking.

A. Organizational Size

A review of the literature pertaining to organization or group size was made. Initial sources were obtained from the extensive bibliography accompanying Volume I of The Managing of Organizations by Bertram Gross. Other articles were found by perusing the Education Index, Psychological Abstracts, Sociological Abstracts, the Science Citation Index, the Review of Educational Research, and the Reader's Guide to Periodical Literature.

The general paucity of articles pertaining to size as a characteristic of organization is illustrated by the fact that only eight articles appeared in the Sociological Abstracts between 1953 and 1963. Similarly, there were only six pertinent articles cited in Psychological Abstracts between 1955 and 1965.

Cross references from the various articles relating to organizational and group size were explored. Twenty-eight were found which were directly related to organizational size, while another twenty were extensively reviewed. In addition, a tabulation of the variables related to organizational size is provided below:

1) Absence (Revans, Indik-C)
2) Accident rate (Revans)
3) Amount of delegation to subordinates (Hanson)
4) Amount of executive responsibility (Hanson)
5) Attitude (Talacchi, Indik-C)
6) Communication problems (Caplow)
7) Compensation (Simon)
8) Creativity (Gibbs)
9) Depersonalization (Burling)
10) Effectiveness (Likert)
11) Efficiency of task performance (Likert)
12) Employee morale (Worthy)
13) Ethical commitments (Thomas)
14) Frequency of succession (Grusky)
15) Group stability (Caplow)
16) Individual behavior (Talacchi)
17) Input (Herbst)
18) Leadership perceptions (Medalia)
19) Management structure (Revans)
20) Member participation (Warner, Indik-A)
21) Morale (Revans)
22) Operating efficiency (Worthy)
23) Organizational environment (Gross)
24) Organizational purpose (Gross)
25) Output (Herbst, Marriott, Revans, Indik-C)
26) Problem-solving ability (Gibbs)
27) Proportion of resources devoted to self-maintenance (Caplow)
28) Punctuality (Revans)
29) Role conception (Thomas)
30) Role consensus (Thomas)
31) Size of administrative component (Anderson, Terriens)
32) Sociological effects on group (Simmel)
33) Strikes of employees (Revans)
34) Supervision ratio (Revans, Indik-B)
35) Tenure of leadership (Kriesberg)
36) Turnover (Indik-C)
37) Uniformity of organizational design (Caplow)
38) Work performance (Thomas)

The literature on organizational size can be grouped into three categories: (1) the effects of organizational size on employees, (2) the relationship between organizational size and organizational output, and (3) the effect of organizational size on the structure of an organization.

In the first category, the size of an organization has been found to have varied effects on employees. In a study of industrial statistics, Revans (1959) found evidence of lowered interest in one's job among employees as organizational size increased. The lower morale was reflected by evidence of greater absenteeism, an increased accident rate, a lower rate of punctuality, and a greater number of employee strikes as organizational size increased. In another study, Talacchi (1960) who studied

British miners, discovered that the rate of subscription among miners to a mining magazine was consistently lower among the workers in the larger mines; the salient conclusion in this research was that the dissatisfaction of workers in the larger mines was significantly correlated with the lower subscription rate to the mining magazine in such mines. In another experiment recently reported by Argyris (1959) it was found that by pursuing a formula of "optimal undermanning", branch banks were able to increase worker productivity, cohesiveness, morale, and cooperation. In a study done by Worthy (1950) Sears-Roebuck employees declined in morale as the size of administrative units and the complexity of the organization increased.

Other studies are also pertinent to the effect of organizational size upon employees. Simon discovered that compensation was a factor positively and primarily related to organizational size. Medalia (1954) in a study of Air Force squadrons found that the squadron leaders were seen as less human persons as the size of the squadron unit increased. In a study investigating the relationship between membership participation and the size of an organization, Warner (1964) and Indik (1961) found that the size of an organization was not inherent to membership participation but rather that a cluster of factors is related to this employee behavior. Finally, in a study of various groups of welfare workers, Thomas (1959) found that smaller groups had persons with higher ethical commitments, greater breadth of role conception, and greater role consensus than did larger groups.

Several studies were found linking organizational size to organizational output. Likert (1961) found that group effectiveness declined above an ascertained optimum group size. In another study, Worthy (1950) discovered that with the increasing size of an administrative component of an organization, worker efficiency decreased. In a study of automobile factory workers, Marriott (1949) found that worker output increased in smaller groups. A higher quality of worker performance was correlated positively with smaller work groups in the study of welfare workers made by Thomas (1959) while a study of college students made by Gibbs (1951) revealed that creativity and problemsolving abilities among students were enhanced by small group organization.

In the review of research regarding the relationship between organizational size and organizational structure, Hanson (1964) found that there was a tendency to delegate more responsibility in larger groups. However, Caplow (1957), Burling (1956), and Argyris (1959) have also found that larger groups are characterized by having greater communication problems and by being more depersonalized. In another study, Kriesberg (1962) found that larger organizations were characterized by a decreased tenure of leadership while Grusky (1962) found that in such organizations there existed a greater frequency of succession in leadership positions. In a study investigating the relationship between the proportion of resources devoted to self-maintenance and organizational size, Caplow (1957) discovered that the proportion of resources allocated to this function increased as the size of the organization grew. A study by Terriens revealed that as organizations increased in size, the administrative component also enlarged. This study is balanced, however, by a study made by Anderson (1961) which showed that the administrative component was proportionately less in larger organizations when other factors such as increased number of tasks to be performed or the increased number of plants in the organization were excluded.

Finally, it is significant to note that there is not a well-defined body of literature on organizational size as it relates to educational enterprises. There are, however, two recent studies whose findings deserve citation. In a study of organizational decentralization in the school systems of Montgomery County, Maryland and Atlanta, Georgia, King (1966) discovered that system decentralization contributed to improved teaching conditions, better communication, and an improved instructional program for students. Also pertinent is the report of the recent study of the New York City School System by Gittel (1967) and Hollander (1964). These investigators maintain that their research data reveal that the New York City School System is virtually paralyzed by its large and ponderous bureaucratic structure. Furthermore, the study asserts that this mammoth organization has produced inertia and hindered the school system from responding effectively to new demands. Evidence for this conclusion is provided by data revealing that the New York City System has not made any meaningful changes for thirty years in curriculum, administrative structure, teacher recruitment, appointment procedures, or training. The investigators suggest that organizational decentralization of the system might contribute to the facilitation of greater organizational effectiveness.

B. School Size

As mentioned above, much of the research which has been done on the influences of school size relates size factors to dollar costs. The resulting evidence has been useful where costs alone are under consideration. Since those studies are not of direct concern in this research, they have not been presented below. The research reported below was found by searching such sources as Dissertation Abstracts, Education Index, Encyclopedia of Educational Research, Review of Educational Research, and Journal of Educational Research.

Attempts have been made recently to relate certain factors of size, especially enrollment, to indices of quality. Quality has been defined in different ways, usually in quantitative terms such as breadth of course offerings, teaching loads, or special services. Some of this research has produced positive correlations between quality, defined in those ways, and size. California (Smith, 1964), North Carolina (Weaver, 1961), Florida (Dungan, 1961), Alabama (Vardman, Crocker, 1958-60), Mississippi (Osborn, 1962), and Missouri (Saville, 1961) have been the loci of studies substantiating the relationship. A number of persons report a positive correlation up to specific points in size; e.g., 1200 students. Beyond this point a plateau is soon reached and negative correlations appear. With further increases in size the benefits did not grow, and the disadvantages of impersonality and lack of contact became evident. Barker and Gump (1964) report that a negative relationship exists between institutional size and individual participation.

Additional studies show that other factors, such as state of residence, parental educational background (Bryant, 1957), social adjustment (Cluff, 1954), and teacher approval (Christensen, 1960), have more effect on the selected measures of quality than does size. The advantages that are cited in the literature regarding secondary schools above 700 enrollment are usually quite specific and may be summarized as follows:

(1) Better faculty
    qualifications
    compensation
    stability

(2) Better conditions
    more efficient administration
    lighter teaching loads
    more resource persons
    lower teacher-pupil ratio
    more materials and equipment

(3) Better programs
    more opportunity to experiment
    more courses
    more extra-curricular activities

The disadvantages cited are usually less specific and more difficult to document: (1) less personal contact between students and teachers; (2) questionable psychological effects on students; (3) lower participation in extra-curricular activities; and (4) weaker staff communication. Most such studies are of schools having fewer than 1000 students.

Information from studies of this kind have been useful in seeking variables to be included in the present research. In the face of the general feeling that large schools may be detrimental in some ways, changes have been proposed which are intended to remedy the situation. "Schools-within-a-school" (Smith, 1961) or "vertical units" (Evans, 1960) with independent faculties and student bodies have been established to try to gain the advantages that come with size, but, obviate the disadvantages that are noted after the "optimum" size has been reached. See Figure 2-B-1.

An analysis was made by Gaumnitz and Tompkins (1950) of survival in high school based on the U.S.O.E. Biennial Survey of 1945-46. The results are in part a function of the variations among states in statutes regarding compulsory attendance. Their findings are as follows:

(1) High schools are not holding over 50% of pupils through to graduation.

(2) The range from highest to lowest states is from 2:1 to 3:1.

(3) Separate school systems for minority groups invariably show lower holding power.

Variable	Source Document
1) Achievement	Bragg, Gray, Jantze, Lamberty, McDaniel, Smith, H. J., Vardeman, Weaver
2) Administration	Dungan, Grady, Garcia, Shelley, Smith, H. J., Watson
3) Communication problems	Evans, Shapiro
4) Curriculum	Andrews, Barnard, Berg, Bragg, Cracker, Dungan, Garcia, Kempston, MacArthur, McDaniel, Osburn, Rougean, Savilla, Smith, C., Smith, H. J., Sollars, Treadway, Vardeman, Woods
5) Extracurricular activities	Andrews, Garcia, Grady, Gray, McDaniel, Weaver, Woods
6) Faculty qualifications	Barnard, Collingsworth, Cracker, Rougean, Vardeman
7) Faculty stability	Gray
8) Faculty training	Weaver
9) Group cooperation	Shapiro
10) Guidance	Dungan, Gray, Rougean, Smith, H.J., Weaver
11) Materials for instruction	Smith, C., Woods
12) Morale	Evans, Zinser, Gaumnitz
13) Personal relations	Smith, H. J.
14) Program	Jackson, MacArthur, Shelley, Treadway
15) Pupil-personnel ratio	Garcia
16) Pupil-teacher relationship	Tyson
17) Range of courses offered	Barnard
18) School-community relations	Andrews
19) Special instruction (ability grouping classes)	Savilla, Weaver
20) Student characteristics	Williams
21) Teacher characteristics	Patterson
22) Teacher effectiveness	Shapiro
23) Teaching of basic skills	Loughridge, McDaniel

Figure 2-B-1

Summary of
Variables Related to High School Size

(4) Schools organized as Junior-Senior High Schools do better on holding than do other types of organizations.

(5) Large high schools are not holding pupils any better than small ones.

Barker and Gump (1964) suggest three alternatives to handle increasing enrollments and preserve small school advantages: (a) establish new small schools; (b) set up campus schools in which students are grouped in semi-autonomous units; (c) set up autonomous campus units. They state that a school should be large enough to have a comprehensive program and small enough so that (i) a higher percentage of its students are needed for its various enterprises and (ii) the enterprises are not redundant. Yet, they conclude that their research does not answer the question, "How large should a school be?" The reason given is that crucial variables related to educational productivity were not investigated.

C. Social Area Analysis

In doing the preparatory work in this study a search was made of reported research to determine the instruments and methods of analysis which would be most useful in the work. Dissertation Abstracts, Economic Abstracts, Sociological Abstracts, and The Educational Index were used as the starting point in assembling pertinent information in regard to the classification of community characteristics. The most useful material found to date is presented below.

"It is a matter of everyday observation," says Wendell Bell, "that metropolitan areas are subdivided into different sections, each exhibiting different features." A number of investigators have tried to capture the distinctive features in some kind of typology that would be useful in analysis.

An approach to a typology will vary with perspective of the investigator. Duncan and Schnore (1959) indicate three different emphases: (1) Cultural, (2) Behavioral, and (3) Ecological. The cultural perspective considers cultural factors as the major predictor of spatial clustering in the city. It emphasizes the symbolic value of space and is critical of the economic bias of ecology. The behavioral or interactionist perspective tends to be concerned with the frequency, intensity and forms of interpersonal communication. The ecological perspective stresses the influences of technology, population characteristics and environment on social organization and spatial distribution. None of these emphases is sufficient to explain the complexities of urban organization, but each concentrates on and elucidates different facets of a multidimensional problem.

Shevky and Bell (1955), in a technique called "Social Area Analysis," acknowledge their indebtedness to the urban ecologist. This social area analysis approach has been used by a number of investigators in their studies of metropolitan areas. The social area is dependent on spatial contiguity.

The basic unit of analysis in the construction of social areas has generally been the census tract. Chicago is divided into more than 1000 of these relatively small geographical areas. Census bulletins contain a fund of information on the population in the tracts--race, sex, marital status, type and condition of dwelling, number of persons per dwelling, income, type of heating, rent, etc.--which is useful in analyzing the area.

Shevky and Bell sought to organize this maze of census data. These data were arranged in socioeconomic, family and ethnic sets. The rationale for this selection is presented schematically in Figure 2-C-1.

Tryon (1955) independently analyzed the San Francisco Bay area and evolved a near identical classification. Tryon employed a more extensive list of variables, 33 in all: eight dealing with population characteristics, thirteen with occupations and twelve with dwelling units. By a method of cluster analysis of the 33 variables, Tryon found seven distinct domains: Socioeconomic achievement, Family life, Assimilation, Female achievement, Socioeconomic achievement 2, Assimilation 2 and Socioeconomic independence.

Tryon gathered his domains under four major factors: Socioeconomic independence, Socioeconomic achievement, Assimilation, and Family life. These empirical dimensions have a validity beyond those of Shevky and Bell, which were the result of theory and insight. The factors are listed below so that the reader can note their similarity.

Shevky and Bell	Tryon
1. Socioeconomic Status (Social Rank)	1. Socioeconomic Achievement
2. Ethnic Status (Segregation)	2. Assimilation
3. Family Status (Urbanization)	3. Family Life

Tryon, after complimenting Shevky and Bell for their insight in choosing on *a priori* grounds what he derived empirically, questioned the fact that they were missing a factor that would parallel his socioeconomic independence. He suggested that their indices would have been improved by using more measures.

Kaufman (1961) revised the Shevky-Bell procedures and tested his new variables by analyzing the census data for the Chicago and San Francisco areas. He dropped rent from the socioeconomic status measure and retained education and occupation, but defined them differently. He also eliminated house type from the family status indicators because of its limited application. The fertility ratio and proportion of women not in the labor force were retained and two additional measures were added, one to qualify the fertility ratio and the other as a measure expressive of family status. The ethnic status measured was redefined as: the proportion who are not native born whites. Kaufman still maintained seven measures but believed that the changes he suggested improved the analysis.

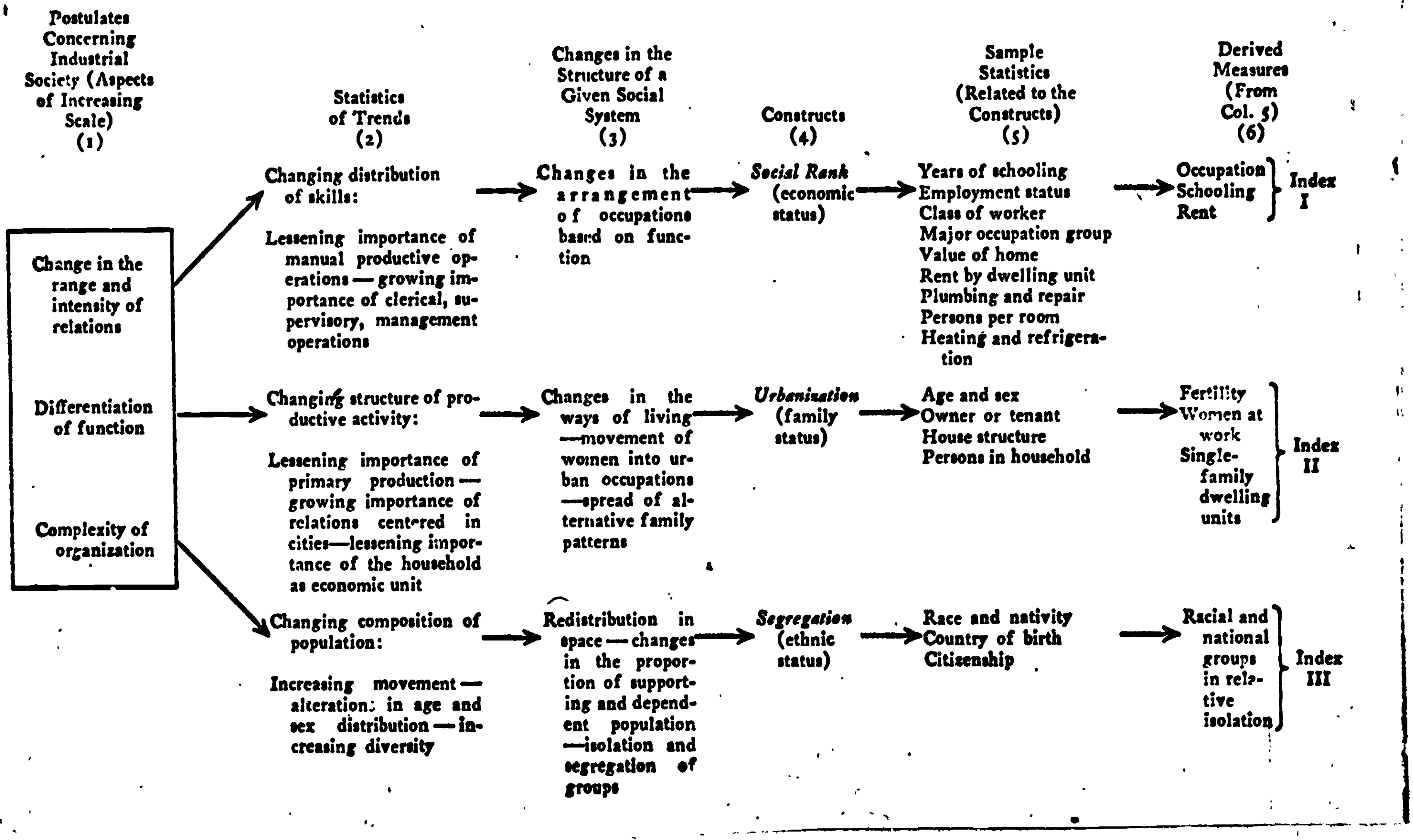


Figure 2-C-1

Steps in Construct Formation and Index Construction*

*Shevky and Bell, 1959, p. 4.

Edward Koch (1966) used an adaptation of the Shevky-Bell analysis in his study of the elementary school districts of Binghamton, Massachusetts. However, instead of dealing with census tracts directly, Koch converted census information to school district data. He also added an index, residential stability, to the three standard Shevky-Bell indexes. This factor, together with social rank, proved to be the most important factor in predicting academic achievement. He found that neither ethnicity nor urbanism was significantly related to academic achievement. Koch's contributions, then, were his method of reducing census tract data to school district data and his addition of a fourth factor, residential stability, to improve the analysis.

Most of the studies reported to date have been devoted to the Shevky-Bell method of social area analysis or assorted modifications of it. The reports which follow are distinct from this model and report on varying geographical areas with many more indicators than any of the previous writers have utilized.

Christen Jonassen (1958) used eighty-two variables in his search for functional unities in the eighty-eight Ohio counties. He used factor analysis, as Tryon used cluster analysis, to see if the vast number of community variables would form independent clusters of highly interrelated components. He wanted to simplify the complex structure of community interrelationships to a manageable list of factors which could account for the differences between community systems.

Jonassen used factor analysis on the eighty-two indicators of Ohio counties and arrived at seven factors or clusters: Urbanism, Welfare, Influx, Poverty, Magni-Complexity, Educational Effort and Proletarianism. Urbanism loaded significantly on fifty-one of eighty-two county measures with very high loadings on clerical and sales workers, urban populations, and social complexity. Welfare loadings were high on efficiency, welfare, educational sacrifice and educational effort. Influx had high loadings on migration gain, population gain and population mobility, among others. Magni-complexity was characterized by high loadings in population density, size, educational plant size, and governmental complexity. Educational effort is self evident; finally, proletarianism had high positive loadings on unskilled workers and a high negative loading on wealth differential.

Byron Munson (1966) at Ohio State University replicated Jonassen's work in Ohio and Charles Bonjean used a majority of the same measures in a national study. Both of these investigators should have reports published on their results soon.

The emergence of these meaningful clusters, Jonassen felt, was an indication of the reliability of the units of measurement and of the basic data. The eighty-two variables used did piece together to yield a significant picture of systems within the established framework of census information.

Roger Lennon (1952) did a study on the "Prediction of Academic Achievement and Intelligence from Community and School System Characteristics." He neatly summarized the studies of Thorndike, Davenport-Rummers, Maller and Pierce and then went on to confirm a number of hypotheses relating achievement and intelligence with a variety of community characteristics. He investigated the extent to which measured achievement of the public elementary school population of a community is predictable from a knowledge of certain social and economic characteristics of the community, and certain characteristics of its school system. He found that median schooling of the adult population, median monthly rent, per cent of adult illiteracy, per cent of native white, and per-pupil expenditure were consistently highly correlated with academic achievement. Achievement is positively associated with qualities pertaining to the intellectual, cultural, and economic status of the adults in the community.

Schutz (1956, 1960) also used the factor analytic approach in two studies, one of educational development in the United States and the other of academic achievement and community characteristics. In both cases he was looking for functional unities which might be useful in further educational research.

The 1956 study of state data was based on thirty variables. Four factors were extracted using Thurstone's complete centroid method. Schutz named the factors urbanization, administrative organization, intellectual climate, technological advances and lagging social institutions.

In his 1960 study, Schutz used twenty variables. The data were collected and correlated by Gawkowski (1955) in his work on the use of community characteristics for obtaining local norms on standardized achievement tests. The variables included indices of the educational background of adults in the community, the labor situation in the community, the racial and national origins of the inhabitants, the financial income and the academic achievement in the community.

Five factors were extracted using Thurstone's complete centroid method: (1) Urban-Financial, (2) Intellectual Climate, (3) Economic Stability, (4) Academic Achievement, and (5) Low Socioeconomic Status. This analysis, based on eighty-four local communities, provides evidence at another geographical level of the existence of these factors. The fact that academic achievement loads only on two achievement variables suggested that high academic achievement can be obtained in widely differing communities.

James Beshers (1956) studied census tract data and social structure. His study yielded four factors from the twenty items of census information which he used. He called his factors social status, young family, female-stability and race.

The work of Johnson (1958) and Kaplan (1958) was directed by Borgatta at New York University. Johnson searched for functional unities in an analysis of United States county data. He found six factors that accounted for 100 per cent variance. Since population size itself accounted for sixty per cent of the data variance, he decided to convert all data not in rate form into per capita rates and to re-analyze. Forty-five significant variables resulting from preliminary analysis were re-analyzed and five factors accounting for sixty per cent of the variance in the second analysis were retained. The factors are: (1) Size of population, (2) Non-farm, urban-rural monetary economy-level of living, (3) High farm-operator family level of living index and stable demographic structure, (4) Age structure of the population, and (5) Residential stability.

Kaplan's analysis of the metropolitan areas of 25,000 or over also yielded five factors which accounted for seventy per cent of the variance: (1) Population size, (2) Socioeconomic dimension, (3) Population stability and growth, (4) Ethnic and racial homogeneity plus residential orientation, and (5) Age-sex structure.

Borgatta directed a similar study by Jeffrey Hadden at the University of Wisconsin (1963). Part of Hadden's dissertation was a replication of Kaplan's work. He then extended his investigation in an effort to determine whether the underlying dimensions that he and Kaplan discovered would obtain when other territorial units were observed. The results of factor analytic studies lent considerable support to the proposition that the major factors discovered represent stable underlying dimensions of urban social structure. These factors remained basically unaltered when different territorial definitions of cities were examined (SMSA's, Urbanized Areas and Cities). The principal value of this empirical approach was that it provided a succinct way of summarizing and ordering the relationships that exist among a large number of variables.

Another study that may be useful in the present investigation is that of Allen and Bentz (1964). They were interested in developing an instrument to measure sociocultural change using state data. They factor analyzed thirty-two indicators which enabled them to assess change in various parts of the society and culture. Their work yielded four factors: (1) standard of living, (2) population growth, (3) industrial-technological-urban development, and (4) education. Since they were dealing with change all their indicators were rates, i.e., percentages. They rated forty-eight states on their indices and were satisfied with their ability to measure change and suggested that future studies may attempt some type of study of causal connections between the components.

Three additional reports have been reviewed. One was a study by Wood (1961) of the public sector of the New York metropolitan region. Included were 22 counties and approximately 1400 governments. In the analysis, seven factors were extracted which were described as "identification tags, not complete descriptions of the contents." They were composed of thirty-four variables, twenty of which were described as socioeconomic and fourteen as fiscal. The seven factors were: (1) community size, (2) industrialization, (3) housing density, (4) age, (5) low income prevalence, (6) residential affluence, and (7) land reserve.

In a study by Peterson, Rossmiller, Wakefield, and North (1963) of a large number of variables associated with ability to support education and need for education support, 161 variables were identified and reduced to fifteen factors. They relate to the following: (1) personal income, (2) certain municipal expenditures, (3) certain municipal receipts, (4) agricultural land valuation, (5) size of certain high school classes, (6) certain school expenditures, and (7) staff-pupil ratios.

The final citation is Reissman's (1964) critique of classification of social data. He analyzed five indices and their indicators. The five indices were: (1) population size, (2) economic function, (3) occupational distribution, (4) moral integration, and (5) social area analysis.

The factors extracted in the fourteen investigations are presented below:

1. Tryon's Clusters (1955)

   Family Life
   Assimilation #2
   Socioeconomic Independence
   Socioeconomic Achievement #2
   Assimilation #1
   Female Achievement

2. Bell-Shevky and Bell (1955-61)

   Social Rank or Socioeconomic Status
   Urbanization or Family Status
   Segregation or Ethnic Status

3. Jonassen (1958)

   Urbanism
   Welfare
   Influx
   Poverty
   Magni-complexity
   Educational Effort
   Proletarianism

4. Schutz -- #1 (1960)

   Urban-Financial
   Intellectual Climate
   Economic Stability
   Academic Achievement
   Low Socio-Economic Status

5. Schutz -- #2 (1956)

   Urbanization
   Administrative Organization
   Intellectual Climate
   Technological Advances and Lagging Social Institutions

6. Allen and Bentz (1964)

   Population Growth
   Industrial-Technological-Urban Development
   Standard of Living
   Education

7. Johnson (1958)

   Population Size
   Level-of-Living
   High Farm Operator Level of Living
   Median Population Age
   Residential Stability
   Commercialized Farming vs. Subsistence Farming

8. Kaplan (1958)

   Population Size
   Socioeconomic Dimension
   Population Stability and Growth
   Ethnic and Racial Homogeneity
   Age-Sex Structure

9. Hadden (1963)

   Population Size
   Socioeconomic Status
   New Residents
   Age of Residents
   South vs. Non-South
   Per Cent Living in Group Quarters
   Population Density
   Per Cent of Employed Worked Outside the County of Residence
   Per Cent Non-White
   Population Increase
   (Nine other factors unnamed)

10. Koch (1966)

    Social Rank
    Residential Stability
    Ethnicity
    Urbanism

11. Kaufman (1961)

Titles same as Bell-Shevky but Indices Differ

12. Wood (1961)

Community Size
Industrialization
Housing Density
Age
Low-income Prevalence
Residential Affluence
Land Reserve

13. Peterson, Rossmiller, Wakefield, and North (1963)

Income
Municipal Revenues
Municipal Expenditures
Equalized Valuation

14. Reissman (1964)

Population Size
Economic Function
Occupational Distribution
Moral Integration
Social Area Analysis

D. Colloquium on the Determination of School-Community Characteristics

On January 5th and 6th, 1967, a colloquium was held at Madison, Wisconsin on the determination of school-community characteristics. Its purpose was to seek additional guidance from special consultants with regard to problems encountered in the conduct of this research. These problems were of three kinds: (1) the dimensions of community relevant in the study of school units, (2) the geographic mapping of such data, and (3) estimation techniques useful in combining data from hierarchical or disconsonant units. The remainder of the colloquium was devoted to discussion and application of the ideas presented on the three problems.

Dimensions of Community Characteristics Relevant in the Study of School Units (Bonjean, 1967)

(An abstract of a formal paper presented by Bonjean is presented below, followed by a summary of the discussion which ensued.)

A major problem associated with comparative analysis, including factor analysis, of communities results from the ambiguity of the concept "community". Previous factor analytical studies have been limited to atypical sets of communities.

The present study utilizes the county as a unit of analysis, and includes variables used in previous studies. The purpose is to locate dimensions which explain variation and note similarities with dimensions identified in previous studies. The county was selected as the unit of analysis because it is the largest administrative unit below the state, it involves both rural and urban situations, and it is as clearly defined politically, socially, economically, culturally and functionally as is the city. Variables were drawn from the U.S. Bureau of Census information and previous studies. The universe of the study is a duplicate, with few exceptions, of the universe used in The County and City Data Book.

The method includes initial factoring by a version of Hotelling's principal axes method to reduce the orthogonal factors, and rotation by normalized varimax procedure to accommodate the factor breadth of the study. In this study variables are called primary variables of a factor if they have their greatest loading on the factor in question.

Eighteen factors accounting for 78.6% of total variance were extracted. The eight accounting for the greatest amount of variance are the following:

I. (accounting for 18.2% of the total variance) The "Socioeconomic Status" factor, which includes variables indicating social stratification, has been the most critical factor in previous studies as well. Variations among the different studies with respect to this factor are most likely due to the inclusion of different indicators.

II. (11.3% of total variance) The second factor includes variables relating mainly to age structure, although considerations of the direction of the loadings and interrelations among variables suggest the name "Family Life Cycle". The non-occurrence of this factor in one previous study might be attributable to the universe of that study since there is considerable overlap in variables and the units of analysis are the same.

III. (6.1% of total variance) Factors similar to this third factor, "Governmental Revenues and Expenditures", do not appear in previous studies, probably due to the fact that most of the important variables clustering about this factor were not used in the earlier studies. The relationship of this factor to factor I is less than expected, although its greater relationship to educational rather than wealth variables might have been anticipated.

IV. (5.6% of total variance) "Residential Mobility" occurs as a factor in previous studies and includes similar variables. This factor shows some relationship to factor 1.

V. (4.2% of total variance) The primary variables associated with "Urbanism" are the classical variables of size, density, and heterogeneity, yet the low percentage of total variance belies the traditional importance of the factor, suggesting a decline in urban-rural differences, a conclusion which can only be tentative, however, for traditional use of these variables, involves parameters not available for the present study. In all investigations, this cluster of variables was found to be relatively independent of the cluster of variables called "Socio-economic Status". Unexpected loadings may be explained by a distinction between classical definitions of urbanism and that used by the U.S. Census Bureau.

VI. (4.0% of total variance) The negative loading of one primary variable in "Manufacturing Concentration" suggests that government centers are not usually manufacturing centers and vice versa. This suggestion is supported by findings of a previous study.

VII. (3.9% of total variance) Another economic specialization factor, "Commercial Center", reveals a limitation of factor analysis. Comparisons of factors of this and previous studies indicate a divergence of results. Attention to detail, however, discloses obscured relationships which usually dissolve the divergence.

VIII. (3.1% of total variance) "Unemployment" does not occur as a factor in previous studies. The critical variable involved, however, does occur in these studies though associated with factors which do not appear in the present study.

None of the remaining factors explains more than 2.9% of total variation.

Comparisons of all the studies considered suggest that the primary dimensions of community may be termed "socio-economic status", "residential mobility", "urbanism", and "non-white". Only two factors which occur in the present study, "government revenue and expenditures" and "unemployment" do not occur in the previous studies considered. However, these factors appear to be the most dependent upon the selection of variables and unit of analysis.

Discussion of Bonjean's presentation: If one is interested in finding independent dimensions, or in clustering out a large amount of data, it is advantageous to use the least number of concepts in an orthogonal analysis, i.e., uncorrelated factors. Perhaps, however, it would be better to start with the assumption that there are independent uncorrelated factors, doing an oblique, and then inferring rather than assuming that these are consistent independent dimensions.

The basis of the 79 variable selection was not theoretical, since a base for selection would draw one away from a pure factor analysis. It was supposed that if too many related variables were used, artificial factors might be created, factors which are a function of the data pulled in rather than of the underlying character of cities. The basis of selection came from a previous study in which the criteria for selecting communities included functional or economic specialization and size. Such selection is not an end in itself -- no factor analysis is -- but it may be useful in one type of description, even though it does not explain anything nor can predictions be made from it.

There seems to be an assumption in this study that large and small counties vary in the same way, for unless one controls for size the other factors do not discriminate.

The studies which have been made allow one to discover relationships among communities. Previous studies have provided clues and rudiments which may allow theory building. A problem involved with any type of social research, however, is drawing boundaries around the social unit concerned. At this time, such boundaries can only be proximate.

Mapping Data and Units (Sale, 1967)

The problem at hand is that of the population of school districts. Such figures are collected on the basis of townships rather than school districts, and the boundaries of townships and school districts are allowed to differ in most states. The approach so far has been developed on the basis of geographic area, assuming an even distribution of population in townships. School district population is to be determined by a sum of township populations lying in the district including any townships which also fall within other districts. The population of these township fractions will be determined by the product of the percent of the area lying within the district in question and the total population of the township excluding any urban areas.

Three rather flat and evenly populated areas of Iowa have been selected for pilot study. One problem with Iowa is the poor map coverage, including school district maps which vary in scale, accuracy and readability. Hence no accurate comparative measurements may be obtained from the maps themselves.

An aim of the pilot study is to determine a method of partitioning data by some kind of geographic division of the units concerned. One possibility is the transference of conclusions from data of one type of geographic area to another type of area. A research project in Texas instigated by Title I of the Elementary and Secondary Education Act revealed that the population in urban areas accounts for much of the school district population.

There may be some difficulty in coordinating the two basic sources of information used in the project, namely the Iowa State Department of Education and the Bureau of the Census. There is the possibility of using the coordinate systems used by the federal census as a basis for analyzing data, but these coordinate systems are planned by the states and manifest some variation although they are becoming more popular for mapping purposes. An attempt has been made to combine school district, civil division, and county maps in order to identify school districts. The results were compared with larger scale maps provided by the counties in question. Estimates and measurements were made from both large and small maps. Considerable variation showed up in some cases.

The possibilities of computerizing data from maps including school district boundaries must wait the technical achievement of transferring the named unit of census school district into a computer. The census data are or will be already in computerized form. The census plans for 1970 include the capability of identifying an area in terms of blocks or block phases which when grouped and tabulated by computer will give information about school districts as well as other districts. The boundaries will be ordered on the basis of computer information, but such information will apply only to urban areas.

A difficulty associated with data other than total population of school districts may be alleviated by taking the school district boundaries to the enumeration boundaries of the census bureau. It has been discovered that school district maps obtained from local school superintendents and the enumeration district maps of the Census Bureau may be successfully combined. The enumeration district, however, limits the selection of variables used, which is a problem only when selected variables are desired. The cost of obtaining enumeration district tapes is estimated roughly at $250,000, exclusive of mapping costs.

There are a number of problems related to projects of this nature. There is, for instance, the question of whether to estimate area percentages or to go to the trouble of measuring them. Also there is a problem of whether to use these roughly constructed maps, or maps from the school superintendents, or state highway maps which plot dwelling units. A problem with converting data to school units is the high flexibility of school unit boundaries. Another difficulty concerns the determination of high school attendance areas. In Florida, for example, this cannot be determined directly by school district population. In Iowa where this may be determined by school districts, the districts cut across township boundaries. Finally, with the county as the unit of analysis, the internal heterogeneity of the unit will be so great that adequate comparative analyses will be impossible unless either smaller units are considered or variables are selected which reflect the differential character of the units.

Estimation Techniques for Hierarchical and Disconsonant Units
(Wiley, 1967)

The Sale presentation dealt with disconsonant units such as school districts and townships with incomparable boundaries and divergent sets of data. The present discussion will be concerned with hierarchical units, units of analysis which are completely contained in other units of analysis. The different levels of these hierarchies will yield data not available at other levels, hence there is a problem of comparability of units. The purpose at hand will be the discussion of some possible solutions to this problem.

The problems of hierarchical and disconsonant units affects the type of research which, on the basis of indirectly available data, attempts to determine variables to characterize a given unit of analysis; in particular, attempts to characterize school districts on the basis of data available only at larger units. Consider the problem of coordinating two sets of data, one available only at the county level, the other only on the school district level. How could differentiating variables for school districts be determined from county data?

One solution to the problem might be to do a regression analysis of the constant county values on the variable district values. Linear dependence may be corrected by mean deviating the predicted values from the predicted county mean and adding back actual county values. The now linear independent scores will still differentiate districts. A second solution, assuming no substantial bias associated with the groupings of districts within a county, would begin by tabulating county variables by cumulation of district characteristics. A regression equation then established on the county level could be used to predict county scores and then fed back to the district level to predict district scores. These predicted values could then be mean deviated and the actual values added back. One might also complete analysis of the covariational structure of a relatively heterogeneous area where complete data are available using random sampling. Equations established on the basis of this area could be applied to other areas. A third solution might be to subject raw county values to a composition program yielding scores on composite variables which will give some optimality characteristics to eventual composite variables.

All of the solutions are inadequate to the extent that they overlook actual social class considerations. All, of course, are solutions only to problems of hierarchical units. Disconsonant units, usually lower level units, may be merged until hierarchical structure begins.

Psuedo-hierarchies might be established by a random grouping of lesser units using variables which would preserve the structure of the lower level, yet by virtue of attenuation would allow generalization back down to the lower level. Non-random groupings may be used only if the created unity in itself is of interest for further study. If different theoretical variables are used at different levels and one then attempts to make comparisons at a low level of the hierarchy, the resulting mass of types of unit may be controlled to some extent due to the effects on lower level variation by the structure at higher levels.

It is possible, finally, to achieve an orthogonal or uncorrelated classification system for the purpose of characterization at each level of a hierarchy by doing separate analyses on each set of data and then doing composites to characterize the units.

Substantive Implications of the Choice of Units and Techniques
(Nasatir, 1967)

Since projects similar to those presented today are likely to set precedents for the solution of school problems, it would be best to assess their proposed procedures for future effectiveness.

There is, among others, a critical problem about the relevance of the data used in these projects to the results desired. The reason for this is an inclination to accept whatever data are available. Yet, the selection of data, to be significant, should have some theoretical basis; there should be some reason for supposing the data used are relevant to the outcomes desired. What is desired is an operative model of causation which would include the problems of feedback and the method. What is done, simple description and redescription, will not produce the variables adequate to such a model.

There is a tendency, compatible with the interests of some data gatherers, to over-emphasize similarities and ignore important individual differences. Generally, data sources are published documentations, which set limits to the possible kinds of recombination. Flexibility, however, is essential to the eventual effectiveness of the system. School district situations are constantly changing, hence rendering data bank material obsolete. It might be better, for example, to substitute definitions in terms of geographic corrdinates for district names. Moreover, one must contend with the structure of variables with which data reporting systems are concerned.

In general, though not in every case, the application of data pertaining to larger units of analysis, e.g., the county, to analyses of smaller, less heterogenous, units is of questionable utility. It is not clear, furthermore, that the information provided by a high level institution such as the U.S. Office of Education is relevant to crucial problems at lower levels. The aim of this project, i.e., the discovery of the impact of the social characteristics upon educational outcome, will be frustrated so long as we continue to accept available documented data without a plausible theoretical basis for selection.

Applications to the Stratification of Schools (Johnson, 1967)

The Central Midwestern Regional Educational Laboratory (CEMREL) exists for the purpose of helping to bridge the gap between research and discovery and diffusion and improvement in the classroom in connection with a region which includes portions of Kentucky, Missouri, Tennessee and Illinois. The Laboratory's concern is to develop research designs especially applicable to educational research and development efforts occurring at the multi-state or regional level. Its primary aim is the development of a stratification of all the sub-units of regions in the fashion discussed in the Wiley presentation. The most promising solution at present involves a modification of the analytic procedures and sampling methodology originally developed by the Wisconsin group. The data considered relevant include "factors" derived from all characteristics which can describe units, and are available in part through the state departments of education and on the national level. It is anticipated that regional stratification on the county level will allow stratification on the district and, eventually, school levels. The research designs developed will be applicable to field projects which will implement curricula through pilot studies in selected schools.

The mass of data, the large number of variables and indices, considered relevant to the CEMREL project suggests that a theoretical basis of selection may be necessary. A counter-suggestion, however, is that the important practical aim of the project is to provide adequate guides at any moment for educational decisions, hence the justification of the stratification developed is not in question, and conceptualization, a long-term scientific process, may be kept at a minimum. The important point is that the data cannot be obsolete if they are to be useful. Data obsolescence might be alleviated, for instance, at the county level by introducing a constant of dating of information and stratification into the system.

Applications to Size and Non-Dollar Costs of Secondary Schools
(Wakefield, 1967)

A specific problem about the comparative costs of expanding an existing high school and building a new one to meet enrollment demands has led to the exploration of the general problem of the relation of quality to size in secondary schools. Quality might be determined, as it usually is, by economic factors, or by so-called non-dollar productivity, which may consist of a set of complex variables.

The present investigation began with the ideal that, given a number of indicators and a notion of the desired productivity in a particular community, one might be able to determine the appropriate high school size under those circumstances. Changing situations and consequent productivity expectations could probably be treated as operational problems. One could, for instance, adjust structural flexibility to population flux prognoses. An unique aspect of the investigation has been a concern for composite variables, or qualities, rather than stratification. The project has been restricted to readily available data due partly to considerations of immediate utility and clientele expectations and partly to discover the utility of existing data in practical decision-making. Two diverse states with extensive available educational data were chosen for study in this first phase of the project.

The methodology consists of the determination of variables, or principal component scores, by interrelational groupings of the original data, factorization of these scores by image analysis and rotations, regression analyses with the image factor scores, and, finally, graphical and tabular summarization of the results.

An advantage of the arrangement of data used may well be the greater possibility of evaluating the reliability of composite variables than of simple variables. The aggregation of items demanded at most levels of the methodology will presuppose well-developed theoretical formulations or commitments which will be capable of testing. The decisions about the selection and naming of variables and factors which determines the outcome of the study must be done on an a priori basis.

CHAPTER 3

PLANS AND METHODS OF INVESTIGATION

In Chapter 1 it was stated that the ultimate aim of research of the type conducted in this report is to create an algorithm for designing schools. Inputs of the algorithm will be the quantitative values an administrator might choose for manipulation, such as the size of the building, and outputs will be the qualitative values of the dependent variables, such as pupil achievement. In essence, the algorithm will enable the educational planner to compute the effects of different combinations of independent variables on the type and degree of transformation induced by his school-to-be.

The school will thus be viewed as a transformer, accepting a variety of inputs and accomplishing various designable transformations before divulging its output.

For any real transformer a tetrad of relevant information can be described: (1) input data, (2) data concerning the transformer itself, (3) output data, and (4) data relating the transformer to its environment. The design of the transformer will consequently be dependent on amassing data which adequately describe the characteristics of the input and environment. The design also requires data describing the output characteristics, for the success of any proposed transformation, by definition, demands a comparison of output and input.

A prerequisite for the study, then, was data which accurately characterized the input, output and environment. However, data alone will not establish the design algorithm; there must be an underlying theory and set of definitions to meaningfully relate the data to categories for design.

Establishing a theory and set of definitions concerning design problems had to be another goal of the research.

The final goal was to generate a mathematical methodology for relating the data to the theory and definitions established. To test the methodology, two states with extensive educational data banks, Iowa and Florida, were initially selected for examination.

In Section A of this chapter the methodology is discussed more completely; in Section B a survey of various environmental factors which summarize the educational climate extant in the 48 continental states is presented; in Section C the types of input and output data available in 24 states are indicated; in Section D the problems encountered in trying to order the Florida data are explained.

A. Sequence of Operations

Three phases of simultaneous research were undertaken. Phases 1 and 2 concerned attempts to related existing input, output, and environment data to a common unit. This process was basically the same as the common-denominator method for relating different fractions to the same base. For this research, the common denominator for all data was defined as the school. The methodology and plans for investigation are summarized in Figure 3-A-1.

The major sources of the data needed were the United States Census Bureau and the state departments of education. The source data required special preparation for analysis; much organization and interpretation had to be made before the structure and substance of the content of the census and educational data banks was apparent. The data stored in the educational data banks were sifted and then indexed in order to facilitate the definitional clustering which was part of the methodology.

The census data had to be related to school districts; the Census Bureau divides communities into sections, the boundaries of which are seldom coterminous with school districts. All data, however, had to be related to the school. Therefore, a method was devised to determine the unit-by-unit correspondences between census and school boundaries.

The third initial phase of project activities involved organizing and indexing relevant results of relevant previous investigations; the concerns of preparing the theoretical framework involved the cream of several different scientific domains. Much of the information gathered for this effort has been summarized and presented in Chapter 2.

B. Survey of Forty-Eight States: Educational Climate

As explained in the introduction to this chapter, Iowa and Florida were selected as the states on which the developed methodology was to be tested. These states were selected because it was believed their educational data banks would allow the most input, output, and environmental data to be examined. However, a discussion limited to just Iowa and Florida might also be limited in application. To determine some of the general aspects of the variation in educational climate in more states a survey of the forty-eight contiguous states was made.

The presentation of the survey which follows is divided into four parts. First, nine theoretically defined constructs hypothesized to be related to educational climate are presented and notes are made on how they would ideally be measured. Second, a clustering of available demongraphic and school data into categories corresponding to the constructs is presented. For each cluster, a composite score was produced for each state; the empirical analysis into composites is shown to lead in some cases to revision of the construct definitions.

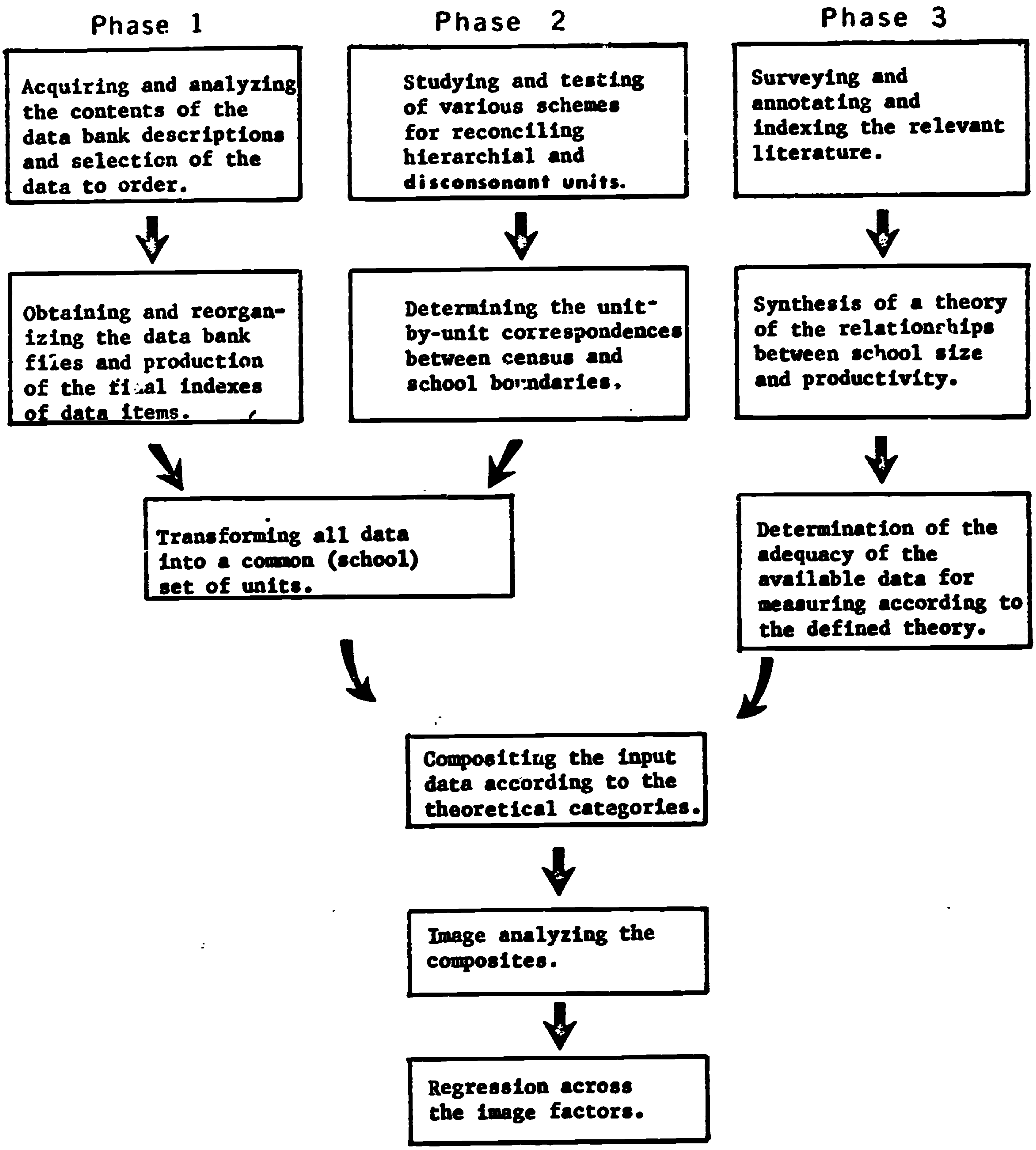


Figure 3-*A-1* *Sequence of project activities*

Third, the factor structure of the composite scores is presented and interpreted. Fourth, the factor scores derived from the factor structure are presented in various forms which affect stratifications of the forty-eight states. These stratifications are used to provide a context for the secondary school qualities of Iowa. (For purpose of completeness, the context of Florida is also considered, although, as explained in Section D below,the Florida data was not analyzed extensively.

Definitions and Ideal Indicators of the Constructs

Nine constructs were defined, partly logically and partly by consideration of possible data sources. Situation, Needs and Innovation, and Motivation and Ability were the three superconstructs from which the nine constructs were developed.

Educational Situations of a state refers to the educational process as it presently operates in the state. Ideally, one would consider measures of pupil input such as family background, natural ability (IQ, etc.) and elementary school achievement. In addition, comparisons of high school curricula and measures of pupil output would be desirable. No data, however, were available in comparable form for all 48 states and four more prosaic measures were used--teacher quality, school size, state control, and funding. Teacher Quality refers to the ability, training and performance of teachers. A measure of teaching performance such as correlation between pupil ability and pupil achievement would be best, but this is not available. A second desirable measure would be a comparison of teacher training such as the number of university credits per teacher. However, this is not obtainable due to the different criteria for teacher accreditation used by each state. School Size refers not only to the number of pupils, but also to measures of departmentalization and specialization. The number of secondary school departments and the number of secondary school courses offered are examples. However, comparative data on these measures has not been collected. State Control refers to state regulation of textbooks, courses, and teacher qualifications. The number of high school courses required by the state and the number of standardized tests employed may represent measures of state regulation. However, there are few such data comparing regular secondary school curricula. Funding refers to the degree of local, state, and federal support for education. A measure of solvency such as the school district indebtedness ratio is desirable. However, the best that can usually be obtained is a breakdown of educational funds into federal, state, and local sources.

Needs and Innovation refers to inferences about the present and future problems of the school system. The expansion and shift of population in a quantitative sense leads to crowding certain areas and relative underpopulation in others. This change involves the destruction of former secondary schools and the building of new ones which are generally larger. During the period of transition, overcrowded and undercrowded schools exist. The qualitative shift of population to urban areas and especially of non-whites to Northern urban areas has more than added to the school population; it has added problems of a lower income group from families of a lower educational level. The opposite situation is the continued existence of secondary schools in population-depleted rural areas with attendance difficulties of maintaining a varied curriculum and a good teaching staff. Two constructs are defined. Quantitative Population Change refers to population expansion in numerical terms; one is interested in ascertaining the potential number of secondary school pupils. Qualitative Population Change refers to the changing composition of the population in terms of racial, urban-rural, and occupational differences. It is best indicated by measuring socio-economic and urban-rural heterogeneity.

Motivation and Ability refers to the desire and ability of the population of the state to improve education. Motivation is best reflected by the community's interest and participation in educational affairs. Support of school bond referenda, percent voting in school board elections, and PTA membership are all examples of community involvement. However, we are not able to obtain this information uniformly and must use two less desirable indicators of motivation: Social Responsibility and Educational Attainment. Social Responsibility refers to the proven interest in group action as reflected by voting, contributions to charity, amount of unionization, and degree of insurance coverage. Educational Attainment refers to the educational level of the community and its reflection in types of occupation. Ability to Pay refers to both wealth and other sources such as non-public schools that are used to defray public educational expenditure.

The constructs are summarized below:

Construct	Definition	Ideal Indicators
1. Teacher Quality	The degree of ability, training and performance of the teaching staff.	1. Correlation between student achievement and pupil ability. 2. Number of university credits per teacher.
2. School Size	The size, departmentalization and specialization of secondary schools.	1. Number of extra-curricular activities. 2. Number of courses offered per secondary school.
3. State Control	Control by state of various aspects of education.	1. Degree of state control over teacher accreditation. 2. Amount of state control over educational curriculum.

Construct	Definition	Ideal Indicators
4. Funding	The money spent on education as derived from state, local, and federal sources.	1. Indebtedness ratio (ratio of school debt per school district valuation) 2. Per-pupil expenditure from federal sources. 3. Per-pupil expenditure from state sources
5. Quantitative Population Change	The change in population through birth and migration and its effect on public secondary school enrollment.	1. Potential number of secondary school pupils. 2. Percent of over-crowded schools.-
6. Qualitative Population Change	The changing composition of the population in terms of racial, urban-rural, and occupational differences.	1. Socio-economic heterogeneity of potential secondary school enrollment. 2. Urban-rural heterogeneity of potential secondary school enrollment.
7. Social Responsibility	Interest and involvement in civic and community affairs.	1. Percent voting for school board members. 2. Percent voting for school board referendum. 3. Percent belonging to PTA.
8. Educational Attainment	The educational level of the population.	1. Median yrs. of education. 2. Status inconsistency.
9. Ability to Pay	The potential monetary resources available to the community for education.	1. Wealth of state 2. Non-public sources of educational funds.

Indication of the Constructs

Direct measurement and, in fact, ideal indication were not possible because of lack of data. Instead, several indicators were gathered for each construct and a composite score was derived by the method of principal components. Data from the period 1958-62 were found to be most readily available and were, therefore, preferred. The titles of the indicators are given in Table 3-B-1 grouped by construct and with principal component loading. The principal component analyses were performed within clusters. In some cases the principal components analysis suggested a revision of the definition of the construct. In the paragraphs which follow, the indicator composition of each cluster is explained and rationalized and any revision of constructs is described.

Six indicators were gathered for the construct, Teacher Quality. These were average teacher's salary, proportion of male teachers, pupil-teacher ratio, supervisors and department heads per pupil, guidance counselors per pupil, and librarians per pupil. The average teacher's salary might be regarded as an indirect indicator of teacher quality since higher pay should attract better qualified teachers. However, salaries are to a large part determined by factors unrelated to teacher quality. The proportion of male teachers may, likewise, be related to teacher quality. Male teachers tend to be younger, are more apt to seek higher degrees and increased specialization, and are more involved in extra-curricular activities, especially in athletics. The pupil-teacher ratio might be expected to correlate highly with teacher quality since smaller classes are felt by many persons to be more effective places in which to teach. The number of supervisors and department heads per pupil, number of guidance counselors per pupil and the number of librarians per pupil are indicative of departmentalization and a more specialized teaching staff. The construct of high teacher quality is positively correlated with average teacher's salary, percent of male teachers, number of department heads and supervisors per pupil, and number of guidance counselors per pupil. It is negatively correlated with pupil-teacher ratio, which means that a low or favorable pupil-teacher ratio is positively correlated with teacher quality. It is also negatively correlated with the number of librarians per pupil. The variation in state accreditation of secondary school librarians and the existence of substandard librarians in many states might account for this negative correlation.

For the second construct, School Size, five indicators were chosen. They include the percent enrollment in small secondary schools, the percent enrollment in medium-sized secondary schools, the change in the number of secondary schools, the number of additional classrooms needed, and the number of excess secondary school pupils. The enrollments in small and in medium-sized secondary schools are a reflection of specialization and departmentalization since larger schools tend to be more specialized and offer a wider range of courses and a greater number of extra-curricular activities. The change in the number of secondary schools is an indication not merely of construction, but also of consolidation. Secondary schools recently built tend to be larger while older schools which are being torn down are mainly smaller ones. The last two indicators, the number of additional rooms needed and the number of excess secondary school pupils, are both indicators of crowding. Crowding often creates school

TABLE 3-B-1

CONSTRUCT INDICATORS AND PRINCIPAL COMPONENT LOADINGS

1. Teacher Quality

+84 Average secondary school teacher's salary
+62 Percent of male teachers
+76 Supervisors per pupil
+84 Guidance counselors per pupil
-42 Pupil-teacher ratio
-64 Librarians per secondary school pupil

2. School Size

-90 Percent enrollment in small secondary schools
+89 Percent enrollment in medium-sized secondary schools
+68 Change in number of secondary schools
-11 Additional classrooms needed
+03 Excess public secondary school pupils

3. State Control

+62 Number of state-required courses
+79 Number of state-recommended courses
+82 Number of state-guided courses

4. Funding

-42 Percent state funds
-70 Percent federal funds
83 Per capita state educational expenditure
08 Percent spent on education
-45 Percent spent on public welfare
93 State educational expenditure per pupil

5. Quantitative Population Change

+84 Population increase
+82 Population mobility
+70 Birthrate
-22 Population under 15 years
+85 Change in public secondary school pupils
+48 Percent of population under 15 years old

6. Qualitative Population Change

+39 Net migration white population
+81 Net migration non-white population
-62 Percent Negro
-51 Change in urbanization
+80 Percent urban
-08 Percent engaged in manufacturing
-09 Percent of technical and professional workers

7. Social Responsibility

+64 Percent voting
+89 Percent covered by hospital insurance
+73 Expenditure for parks
+78 Amount raised by community chest
+87 Membership in AFL-CIO

8. Educational Attainment

+89 Percent with 4 years of high school
+82 Percent with 4 years of college
+90 Median school years
-69 Percent illiterate
+71 Number of doctors
+77 Number of psychologists
+85 Number of technical and professional workers
+59 High school graduates

9. Ability to Pay

+92 Per capita income
+24 Change in per capita income
-93 Poverty
+71 Percent in non-public secondary school
-08 Change in non-public secondary school enrollment

shifts and adversely affects many of the benefits of large schools, e.g., extra-curricular activities. The construct which is determined through the indicator intercorrelations is larger school size. It correlates positively with the percent enrollment in medium-sized schools and the change in the number of public secondary schools. It is negatively correlated with percent enrollment in small secondary schools. It is not related to either of the crowding indicators and, therefore, the construct is that of size rather than utilization of facilities.

The third construct, State Control, consists of three indicators -- number of state-required courses, number of state-recommended courses, and number of state-guided courses. Unfortunately, no comparative data are available on state control over the established secondary school curricula. The indicators are for state control over new courses such as driver education and conservation which were innovated as of December, 1955. No new data were available on a comparative basis and reliance on old data of a more restricted nature would have limited the value of these indicators. In the absence, then, of better indicators these are used. The indicator intercorrelations are high and the construct measures state control over education.

The fourth construct, Funding, consists of six indicators - percent state funds, percent federal funds, per-capita state educational expenditure, per-pupil state educational expenditure, percent spent on education, and percent spent on public welfare. Both per-capita and per-pupil state educational expenditure reflect the actual amount spent by each state on education. In addition, the per-pupil state educational expenditure when compared to the per-capita expenditure takes into consideration the burden borne by non-public education. The percent of educational funds from federal and state sources is related to solvency. Greater federal aid is associated with need and is indicative of a lack of resources from state and local sources. The percent spent on education and the percent spent on public welfare are related to sacrifice and the willingness of a poorer state to divert more state funds into education. The construct turns out to be educational expenditure and correlates positively with per-pupil and per-capita state educational expenditure and negatively with percent federal funds and to a lesser extent with percent state funds. This indicates that in states with high educational expenditure a large part of the funds comes from local sources. There is no correlation between the construct educational expenditure and the percent of state funds spent on education and there is a small negative correlation with the percent spent on public welfare.

There are six indicators of the construct, Quantitative Population Change. They are population mobility, percent of population under 15 years of age, birthrate, and change in number of secondary school pupils. All these indicators reflect the potential secondary school population and its location. All are positively correlated with the construct measuring population increase.

There are seven indicators of Qualitative Population Change. These include net migration of white and non-white population, percent Negro, percent urban, change in urbanization, percent engaged in manufacturing, and percent who are technical and professional workers. The net migration of white and non-white population indicates a differential population change which is reflected in different educational needs. This is likewise true of the percent urban and change in urbanization. The percent Negro, percent engaged in manufacturing and percent who are professional and technical workers reflect the heterogeneity of the population. The construct correlates positively with net migration of non-white population, percent urban, and to a lesser extent net migration of white population. It is negatively correlated with percent Negro and change in urbanization. The revised construct is, therefore, found to be a measure of urbanization.

The construct, Social Responsibility, consists of five indicators - percent voting, percent covered by hospital insurance, expenditure for parks, amount raised by community chest, and membership in AFL-CIO. A more civic-oriented and socially responsible community as measured by the above indicators would be more likely to show interest in its educational system. The indicator intercorrelations for this construct indicate a high positive correlation with all of the above indicators.

The construct, Educational Attainment, consists of eight indicators divided among educational level and occupational level of the population. These include percent with four years of high school, percent with four years of college, median school years, percent illiterate, number of doctors, number of psychologists, percent of professional and technical workers, and number of high school graduates. Educational Attainment may reflect community interest and involvement in educational affairs since a more highly educated and literate population will more likely strive for better education. The indicators have high intercorrelations (positive, except for percent illiterate) and the construct measures attainment.

The final construct, Ability to Pay, consists of five indicators: per-capita income, change in per-capita income, poverty, percent attending non-public secondary schools, and change in non-public school attendance. The per-capita income, change in per-capita income, and poverty indicate the wealth of the community which imposes limits on the funds available for education. The percent in non-public secondary schools and change in non-public secondary school enrollment reflect a source which takes away some of the burden of public education. The indicator intercorrelations are highly positive for per-capita income and percent enrollment in non-public schools, and highly negative for poverty. The construct measures the ability to the state to pay for its educational system.

Factor Structure of the Composite Scores

The composite scores for the clusters are to be considered more fundamental than the indicators on which they are based. But they are considered to vary in a multivariate interrelationship. Therefore, a principal component analysis was performed on the nine composite scores. The intercorrelations, principal component loadings, and varimax loadings are presented in Table 3-D-2. To determine the varimax loadings, the

TABLE 3-B-2

INTERCORRELATIONS AND FACTOR STRUCTURE OF THE COMPOSITES

Correlations

	1	2	3	4	5	6	7	8	9
1	100	60	02	+78	04	78	68	77	91
2	60	100	05	+45	24	61	50	63	64
3	02	05	100	+15	-01	11	27	18	16
4	-78	-45	-15	100	-04	-62	-74	-78	-81
5	04	24	-01	+04	100	01	-30	21	04
6	78	61	11	+62	01	100	65	69	80
7	68	50	27	+74	-30	65	100	62	78
8	77	63	18	+78	21	69	62	100	81
9	91	64	16	+81	04	80	78	81	100

	Principal Components	Rotated Components		
1. Teacher Quality	+92	93	-01	-07
2. School Size	+72	73	31	-04
3. State Control	+19	07	-00	99
4. Funding	87	+86	-05	+13
5. Quantitative Population Change	+06	04	97	01
6. Qualitative Population Change	+85	86	-01	01
7. Social Responsibility	+82	81	-39	24
8. Educational Attainment	+88	87	21	14
9. Ability to Pay	+96	95	-01	09

three unrotated components corresponding to latent roots of the correlation matrix which were greater than or equal to 1.0 were rotated. The principal component and the three rotated components are described in the following two paragraphs.

The principal component, as noted on Table 3-B-2, correlates positively with all but composites 3 and 5. There is a high positive correlation with Teacher Quality, School Size, Funding, Qualitative Population Change or Urbanization, Social Responsibility, Educational Attainment, and Ability to Pay. A high score would seem to imply a high quality of secondary education both in terms of ability and of performance. There was no correlation between the principal component and the constructs of State Control and Quantitative Population Change.

The first rotated component is essentially the same as the principal component, and might be summarily termed Quality of Secondary Education. The second unrotated component consists essentially of composite 5, Quantitative Population Change or Growth. A high score on the factor should imply a great increase in population and a consequent need for change in the secondary school system. The third rotated component consists essentially of composite 3, State Control. As mentioned above, there were no data available on a comparative basis for regular secondary school curricula. The data used were old and were related to special recently-introduced courses such as conservation, atomic energy, and driver's education. As a result, this third component was not used since it did not give a complete and accurate picture of state control over education.

Stratification in the Context of Iowa and Florida

The scores for each state corresponding to the principal component and to the three rotated components were computed. These scores provide a means of stratifying the 48 states. The first stratification is non-dimensional and consists of the ordered list of principal component scores. A diagram containing the ordering is presented in Figure 3-B-1. The divisions between north and south and between urban and rural with regard to educational quality are noteworthy. Moreover, the highest educational quality is found in big-city states from the urban north, followed by the urban midwest and west, the rural midwest and west (including the rural northeastern states of Maine and Vermont), and finally the predominantly rural south and border states. Of the ten states with the lowest educational quality, eight are southern and two are border states.

As previously explained, the third rotated component was difficult to interpret. The data on which it was based were older and of a special nature. However, the second factor is clearly interpretable as "coming needs" and a bivariate stratification with rotated factor 1 provides a more detailed stratification of the states. This is presented in Figure 3-B-2. The second factor, Population Growth, interpreted in our context as "coming needs" does not fit into any regional or urban-rural pattern as does the first factor Educational Quality. Examination of the graph in Figure 3-B-2 reveals that one quadrant is virtually empty.

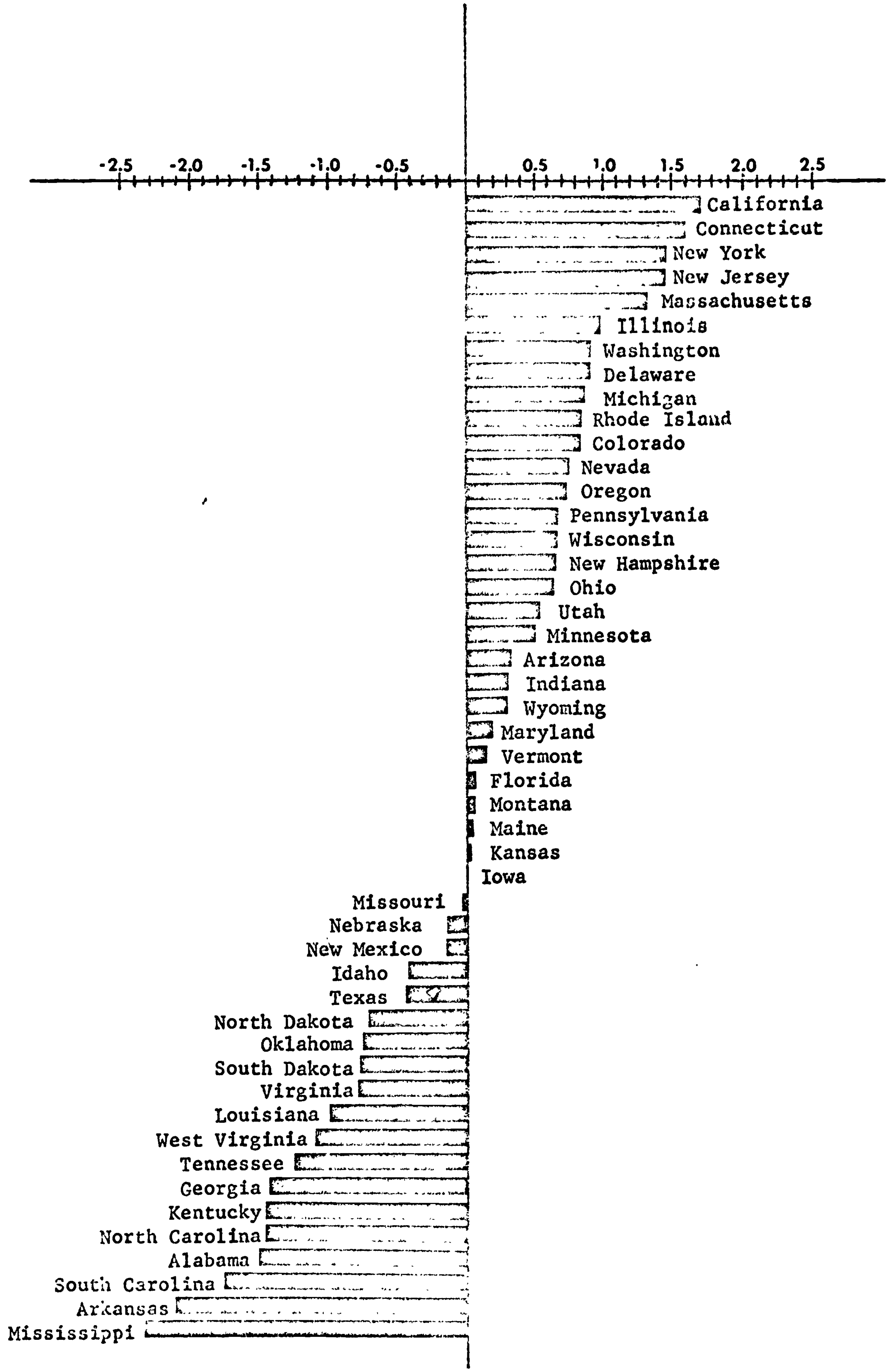


Figure 3-B-1

Stratification Based on the Principal Component

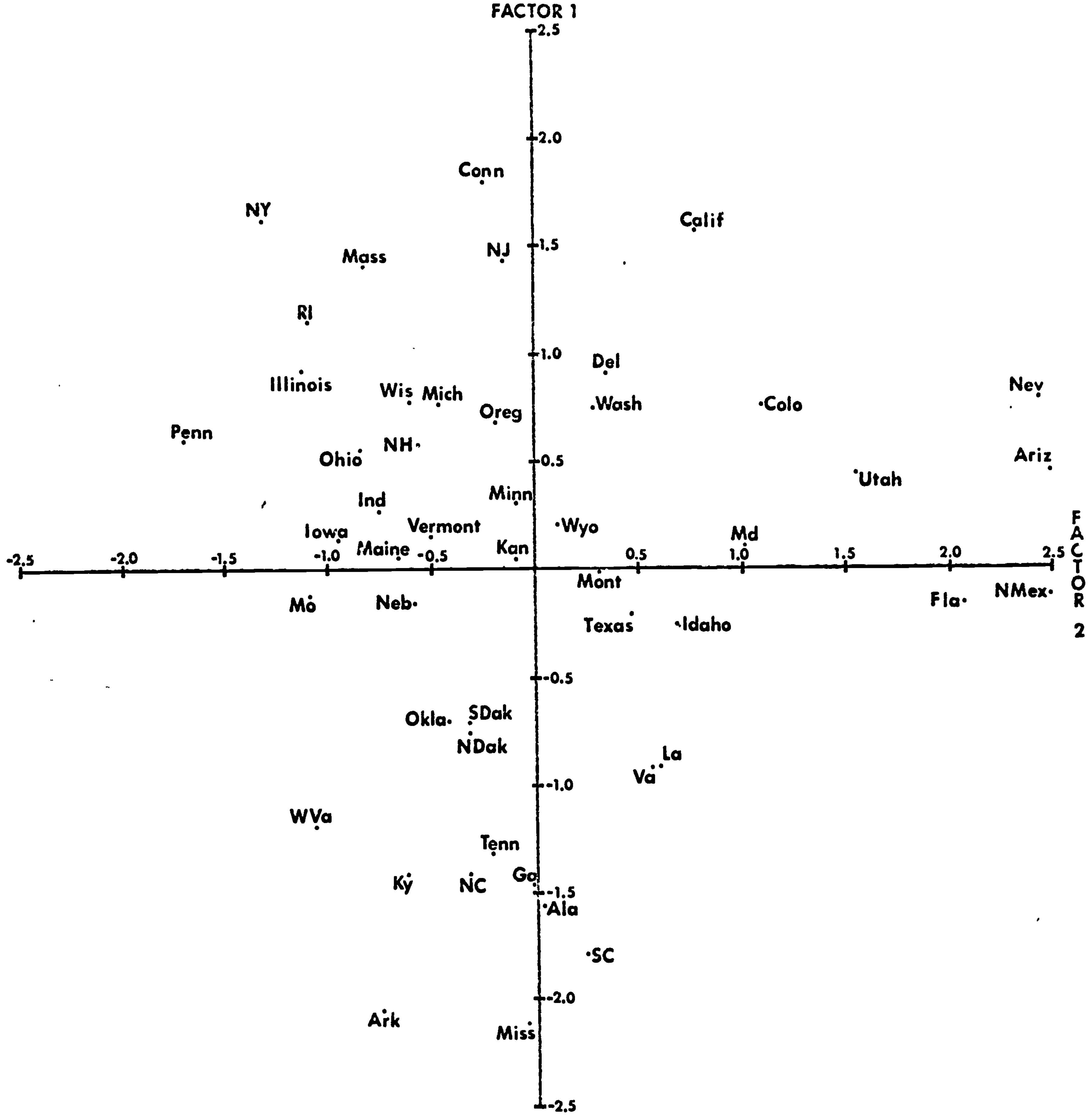


Figure 3-B-2

Stratification Based on Rotated Components 1 and 2

(1st rotated component: quality of secondary education
2nd rotated component: quantitative population change)

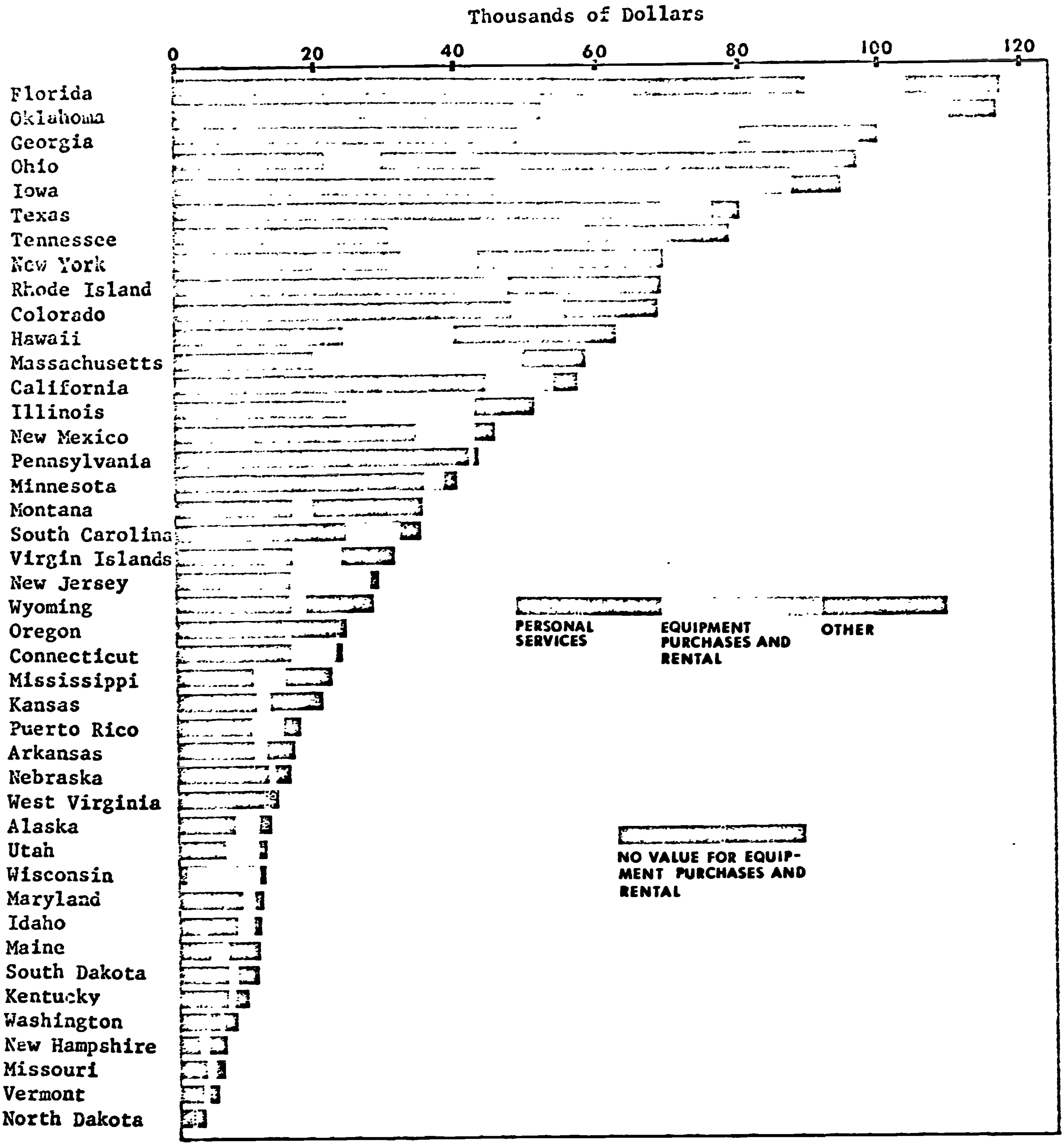
Practically no states which have had a rapid population growth also have very poor secondary school systems. Growing areas almost all have average or better education in terms of the constructs representing ability and performance. Iowa and Florida are about in the middle and are together in the univariate stratification on general quality (Figure 3-B-1). Iowa's position with respect to Educational Quality is similar to other rural midwestern states, while Florida shows a better quality of education than other southern states. With regard to Population Growth and "coming needs" the two states are far apart (Figure 3-B-2). Again Iowa's position is similar to other states of the rural midwest while Florida's phenomenal growth is much greater than that of any other southern state. With regard to the nation as a whole, Florida and Iowa tend toward the middle with respect to Educational Quality. In regard to "coming needs" those of Florida well exceed those of the nation as a whole while Iowa's are substantially below the median.

This analysis suggests that Iowa and Florida do not represent a wide range of educational quality; they do represent states with a great difference in population growth and, therefore, different emerging needs and problems. This condition relates directly to the problem of high school building and size. In Florida where a large building program is in progress, the optimum size of a high school is of crucial importance. In Iowa the problem is one of retaining good education in areas of population stability. The methodology developed in this project is intended to shed light on how both problems can be handled.

C. Automation of Educational Data

The educational data--i.e., the data concerning pupils, teachers, schools, and school districts--were to be obtained for this research from the state departments of education in Iowa and Florida. The project was restricted to the use of data already available in machine-readable form and was, in part, a test of the utility of such data in providing educationally meaningful research results. An evaluation of the utility of the data obtained from the Iowa Department of Public Instruction is presented in Chapter 5. A parallel line of interest concerns the degree to which such data are available throughout the nation. It is important to the eventual extension of the present research to ask where else such data are available. Two presentations are made in this section which are intended to answer that question.

To encourage states to collect and store educational data, Congress included in the National Defense Education Act (NDEA) of 1958 a section providing funds to state departments of education for improving their statistical services (Section 1009, Title X). Upon approval of a proposal for collecting such data, a state could receive up to $50,000 in each fiscal year, provided the federal funds were matched. Figure 3-C-1 is found in "State Plans for Improving Statistical Services," U.S. Department of Health, Education and Welfare, OE-20028, Bulletin 1961, No. 16. It indicates the dollars allocated for educational automation in fiscal year 1960 and thus the relative degrees to which the state departments of education were at that time moving toward the storage of educational data. States which are high on the list may be expected to have rather complete data. Iowa and Florida are high on the list and this corresponds to the fact that they both now have extensive data banks.



Amount of expenditures under section 1009, title X, by State and object: Fiscal year 1960.

Figure 3-C-1

An Indicator of Educational Automation, 1960

To obtain more precise and up-to-date information on the availability of educational data, a telephone survey of 22 states was made in August, 1967. The questionnaire is shown as Figure 3-C-2. A call was made to the director of data processing in each state or to an individual possessing a similar title. He was requested to indicate which secondary education data has been stored in a central facility. He was also asked whether the data were stored in machine-readable form--i.e., on punch cards or magnetic computer tape. Items of the data in the survey include typical information about pupils, teachers, schools and school districts.

A tabulation of the results of the survey is presented in Table 3-C-1. The tabulation includes information about Iowa and Florida; although they were not included in the telephone survey, the information about them was known from earlier project work. The 22 states in the telephone survey were selected at random from the 48 contiguous United States minus Iowa and Florida. The figures in Table 3-C-1 should therefore be a good representation of the status of educational automation across the entire United States. A preliminary observation is that when data are stored, they are usually stored in machine-readable form.

A summary of the results appears as Table 3-C-2. This summary gives for each data item the number of states that have complete data in machine-readable form. Such data could be input to the kinds of analysis used in this research. It is clear from the table that many states have detailed information on teachers and many have enrollment and financial figures for schools or school districts. Several states have curricular information on schools. Only one state (Iowa) has detailed information on individual pupils. This progression is predictable on the basis of the purposes for which the state data banks have been established; namely, the state departments have established data banks in the course of ordinary data processing needs. Since state departments are usually responsible for teacher certification, they need detailed teacher information; since they are responsible for accreditation of schools and for distribution of state funds, they need enrollment, financial, and perhaps curricular information. With the exception of statewide pupil testing programs, state departments are usually not responsible for processing information about individual pupils, so they have had no occasion to include such information in their data banks. Moreover, in many states the compilation of individual pupil records at the state level is viewed with considerable apprehension.

In summary, although data banks are being built, the information now stored by the 24 states which were surveyed (except for Iowa) is not sufficient for the analysis used in this research, since pupil productivity is a major dependent variable complex and requires detailed pupil information. Furthermore, it seems unlikely that pupil data will soon be available in many states. That development awaits increased service involvement by the states in the local processes of scheduling, maintaining transcripts, and reporting grades.

USOE 1266

Survey of Twenty-four States: Education Automation

1. The University of Wisconsin is engaged in a detailed analysis of secondary school size in the state of Iowa. In order to understand the context of the Iowa data, we are surveying other states about how much secondary education data is stored in a central location and about whether data are stored in machine-readable form.

2. This is merely a survey: no request for access to the data is anticipated.

3. Here is a list of secondary school data items which are stored by some states. We should like to know whether your state does store these data (by school) in a central location. If the answer is yes, we should like to know whether the data are stored for all public secondary schools (or teachers, or pupils), or for just some. We should also like to know whether the item is stored in machine-readable form--cards or magnetic computer tape.

4. Here is the list:

	STORED			MACHINE-READABLE	
	not at all	for some	for all	Yes	No
a. Pupil grades					
b. Pupil test scores					
c. Pupils' family backgrounds					
d. Pupils' plans and aspirations					
e. Teacher salaries					
f. Teachers' education and background					
g. Teacher years of experience					
h. Teacher assignments					
i. Curriculum offerings					
j. Number of pupils					
k. Number of teachers					
l. Expenditures					

Figure 3-C-2

Questionnaire on Education Automation

TABLE 3-C-1

RESULTS OF TWENTY-FOUR STATE SURVEY, INCLUDING IOWA AND FLORIDA, AUGUST, 1967

	Stored			Machine-Readable	
	for all	for some	not at all	Yes	No
Pupil grades	1	2	21	2	1
Pupil test scores	2	8	14	7	3
Pupils' family	1	4	19	4	1
Pupils' plans and aspirations	1	4	19	4	1
Teacher salaries	23	0	1	22	1
Teachers' education and background	18	3	3	20	1
Teacher years of experience	18	1	5	18	1
Teacher assignments	17	2	5	18	1
Curriculum offerings	8	2	14	8	2
Number of pupils	20	1	3	19	2
Number of teachers	21	2	1	22	1
Expenditures	20	1	3	15	6

TABLE 3-C-2

SUMMARY OF TWENTY-FOUR STATES HAVING COMPLETE DATA IN MACHINE-READABLE FORM, AUGUST, 1967

Category of Educational Data	Number of States Having Complete Data in Machine-Readable form.
Teacher Salary	22
Number of Teachers	20
Number of Students	19
Teacher Educational Background	18
Teacher Experience	18
Teacher Assignments	17
Educational Expenditure	16
Curriculum Offering	8
Pupil Test Scores	2
Pupil Grades	1
Pupil Family Backgrounds	1
Pupil Plans and Aspirations	1

D. The Data in Florida

The Florida State Department of Education has for the last several years been preparing an information system for education. A state-wide compilation of educational statistics has been made and, as in Iowa, most of the data have been stored in machine-readable form. As mentioned earlier, it was precisely the extensiveness of data storage that led to the selection of Iowa and Florida as the locales for this research. They also seemed to provide contrasting demographic characteristics and educational administrative structures.

The Florida information system has been especially directed at teacher and school certification and accreditation. Therefore, the data bank contains extensive information on teacher background and assignments, on school district enrollment and finance, and on school curriculum. In the course of this research, however, it was learned that Florida data do not include information about individual pupils. Specifically, of the 361 secondary schools, 43 have pupil data included in the data bank. These 43 schools participate in the System for Processing Educational Data Electronically (SPEDE) which enables school schedules, pupil grades, and transcripts to be processed and stored in a central facility. For the SPEDE schools, the pupil data are as extensive as in Iowa, but for the non-SPEDE schools, there are no individual pupil data available.

The lack of complete pupil data in Florida necessitated a halt in the Florida analysis, while the original plans and methods of investigation continued to be applied to Iowa. It was considered of interest, however, to compare the general characteristics of the 43 SPEDE schools with the 318 other schools. The results of this comparison established that the SPEDE schools could not be regarded as a representative sample of the population and, consequently, that analysis of them could not have been generalized to the entire state.

In Table 3-D-1 a list is given of 32 variables identified in the 67 counties of Florida. The variables were extracted from the U.S. Census Bureau City/County Data Book (1962). The 32 x 32 correlation matrix of the variables was computed on the basis of the 67 counties. It was factored according to Harris's (1962) modification of Guttman's (1953) image analysis, and the factors were rotated according to Kaiser's (1958) normal varimax orthogonal procedure. The loadings on the five rotated factors which accounted for the most variance appear to the right of the variables in Table 3-D-1.

The first factor is readily identifiable as Urbanization; the urban counties in Florida are the wealthy, recently-developed areas. The second factor correlates with per-capita and per-pupil local school expenditures, and so is called Educational Expenditure. The third factor is called Youth since it correlates positively with the birth rate and negatively with median age. The fourth factor, Trade Growth, indexes sales and trade expansion. The fifth factor, Public Wealth, correlates with wholesome financial circumstances.

TABLE 3-D-1

FACTOR MATRIX FOR FLORIDA DEMOGRAPHIC VARIABLES

		I	II	III	IV	V
1.	percent urban residence	58	*	*	-30	-25
2.	percent employed in agriculture	-47	*	*	21	*
3.	percent in one-unit houses	-30	20	*	30	*
4.	median population age	29	*	-89	*	*
5.	median school year completed (adults)	86	*	*	-22	*
6.	percent employed in white-colar occupation	77	*	*	-21	*
7.	percent in labor force (adults)	*	*	*	*	*
8.	percent voted in 1960 election	*	*	-27	*	*
9.	percent of vote for leading party	-47	*	*	35	46
10.	percent of vote for Democrats	-64	*	45	*	27
11.	percent non-white	-45	*	21	*	*
12.	percent sound structures	86	*	-24	*	-20
13.	percent overcrowded houses	-48	*	54	*	*
14.	population change 1950-1960	85	*	*	*	*
15.	population density	26	*	-23	*	*
16.	in-migration rate	89	*	*	*	*
17.	birth rate	21	*	87	*	*
18.	per-pupil educational expenditure	*	96	*	*	*
19.	per-capita educational expenditure	*	97	*	*	*
20.	per-pupil proterty tax revenue	-25	22	*	*	65
21.	per-capita total local expenditure	*	66	*	*	30
22.	expenditure-revenue ratio	*	*	*	*	*
23.	retail sales volume	-23	*	*	90	*
24.	revenue-indebtedness ratio	*	*	*	21	83
25.	percent preschool population	*	*	94	*	*
26.	percent recent move to new home	89	*	*	*	*
27.	percent elementary school population	-61	*	61	*	*
28.	marriage rate	*	*	*	23	*
29.	median family income	81	*	23	*	*
30.	trade growth 1950-1960	-23	*	*	91	*
31.	number of unemployed persons	*	*	*	*	*
32.	percent employed outside county	-37	*	*	*	21

Note: Entries between -19 and 19 are printed as "*", and decimal places have been omitted.

TABLE 3-D-2

COMPARISON OF COUNTY FACTORS WITH SCHOOL TYPE

	SPEDE	non-SPEDE	
Factor 1: Urbanization			$X^2 = 3.8$ (N.S.)
High Urbanization	22	212	
Low Urbanization	21	106	
Factor 2: Educational Expenditure			$X^2 = 8.2$ ($p < .005$)
High Expenditure	16	56	
Low Expenditure	27	262	
Factor 3: Youth			$X^2 = 0.4$ (N.S.)
High Youth	18	151	
Low Youth	25	167	
Factor 4: Trade Growth			$X^2 = 3.2$ (N.S.)
High Growth	17	82	
Low Growth	26	236	
Factor 5: Public Wealth			$X^2 = 3.4$ (N.S.)
High Wealth	8	106	
Low Wealth	35	212	

In order to evaluate the distribution of the SPEDE schools, stratifications of the counties were derived from the factors. For each factor, the counties were divided at the median into "high" and "low". The tabulations appearing in Table 3-D-2 were then constructed. For each factor, the corresponding tabulation gives the number of SPEDE and non-SPEDE schools appearing in counties above and below the median. In addition, a chi-square was computed to test whether SPEDE and non-SPEDE schools are distributed differently across the factor.

The chi-square for the second factor, Educational Expenditure, is quite significant. This means that if a school is located in a county which has had a high rate of educational expenditure, then it is more likely to be a SPEDE school. This suggests that SPEDE schools have special qualities, for they have a pre-SPEDE history of high educational expenditure--i.e., of manifest concern and willingness for improving secondary education. An alternative interpretation is that areas with high expenditure have been building new schools, and administrative innovation is easier to effect in new schools. In either case, the differences doubtless correlate with factors such as teacher quality and community involvement. The SPEDE schools may not be considered representative of the population of schools in the state. Therefore, analysis of them alone is inappropriate in terms of the purposes of the present research.

The results of this analysis suggest that a stratified sampling methodology may well be useful in extending to a national scale the present research into high school size and educational productivity. The choice in such extension is between selecting schools where machine-readable data are available, thus greatly restricting the generalizability of the results, and sampling schools according to a stratification plan, thus necessitating the generation of raw data.

CHAPTER 4

DETERMINATION AND PREPARATION OF THE DATA

Modifications of a research plan are to be expected but are not always reported in sufficient detail to be useful in subsequent efforts. In the course of acquiring and processing data in this research project, several unanticipated developments appeared which called for such modifications.

Section A contains an account of contacts and relationships established in the search for data. In Section B is an account of problems which arose in the processing of educational data. Section C is devoted to the problems encountered in the processing of census data.

A. Searching for Data

The search for data was in essence a search for persons sufficiently close to the data who could provide substantive answers to questions of data availability and format. However, they needed to rank sufficiently high in the administrative structure of the system to reduce the number of persons involved in obtaining access to the data.

In each of the two states, official contact was initially made with the Superintendent of Public Education. With his approval, the person identified by him as responsible for data processing was contacted. Through this latter person, a member of the data processing staff was assigned to be the liaison with the research staff.

In addition, a separate relationship was established with the U.S. Bureau of the Census. Letters were written to professors of sociology in each state asking for information regarding any recent analysis or compilation of relevant socio-economic data in that state.

Several persons were retained as continuing consultants to the project. These persons had no direct association with either state. They were called upon as the need to settle either substantive or methodological questions arose. Tables 4-A-1 and 2 list the persons contacted and the data manuals obtained.

TABLE 4-A-1

PERSONS CONTACTED IN THE SEARCH FOR DATA

I. State liaison persons

A. Iowa:

1. Dr. Ralph Van Dusseldorp, Director
Iowa Educational Information Center

2. Dr. E. Gordon Richardson, Assistant Director
Iowa Educational Information Center

3. D. J. Gilliland, Reorganization Consultant
Iowa Department of Public Instruction

4. Professor Lyle Shannon, Chairman
Department of Sociology and Anthropology
The University of Iowa

5. Professor John Hartman
Department of Rural Sociology
Iowa State University

B. Florida:

1. Robert Sims, Systems Coordinator
Florida Department of Education

2. Everett Yarbrough, Systems Coordinator
Florida Department of Education

3. Dr. Archie Johnston, Systems Analyst
Florida Department of Education

4. Professor T. Lynn Smith
Department of Sociology
University of Florida

II. National-Regional liaison persons and continuing consultants

A. Continuing consultants:

1. Dr. David E. Wiley
Professor of Education
University of Chicago

2. Dr. David Nasitir
Professor Sociology
University of California

TABLE 4-A-1 (cont'd)

B. National-Regional liaison persons:

1. Dr. Thomas Johnson, Assistant Director
   CEMREL
   St. Ann, Missouri

2. Dr. Charles E. Hall, Director of School Studies
   Project Talent

3. Dr. Charles Bonjean
   Professor of Sociology
   University of Texas

4. Herman P. Miller, Chief
   Population Division
   U.S. Bureau of the Census

5. Levis J. Conger
   U.S. Bureau of the Census

6. Marshall Turner
   U.S. Bureau of the Census

7. Dr. Byron Munson, Director
   Center for Community and Regional Analysis
   The Ohio State University

8. Arnold A. Heyl, Director
   Division of Data Sources and Standards
   U.S. Office of Education

9. Dr. Richard Powers, Chief
   Operations Analysis Division
   Economics Branch, U.S.O.E.

TABLE 4-A-2

DATA MANUALS OBTAINED

---

1. Iowa Format Book
   content: file documentation for the thirty-eight files (explained fully in Section B of this chapter).

2. Iowa Educational Data Bank User's Manual
   content: data information revealing some items of available data.

3. Iowa CardPac Administration Manual
   content: instructions to principals and proctors for administering a set of student information forms.

4. Instructions for Completing the Pupil Inventory
   content: instructions to students for completing a CardPac questionnaire.

5. Iowa CardPac Student Questionnaire Summaries 1965
   content: statewide summaries of students responses to a questionnaire.

6. Instructions for the Iowa Professional Employees Data Sheet
   content: instructions for completing a teacher information form.

7. Iowa Educational Directory 1965-66 School Year
   content: listing of districts, schools, and staff including identifying code numbers.

8. U.S. Census of Housing: 1960
   Availability of Published and Unpublished Data
   content: information regarding availability of data items per census unit.

9. U.S. Census of Population: 1960
   Availability of Published and Unpublished Data
   content: information regarding availability of data items per census unit.

10. U.S. Census of Population: 1960 Final Report PC(1)-17B
General Population Characteristics for Iowa
content: published tables for the large Iowa census units giving data on age, race, household relationship, marital status.

11. U.S. Census of Population and Housing: 1960
Geographic Identification Code Scheme for Iowa PHC(2)-17
content: names and code numbers for Iowa census units.

12. U.S. Census of Population and Housing: 1960
PHC(1) Final Report for the five SMSA cities of Iowa
content of each of the five reports: published tables of data for census tracts, map of the SMSA with tract lines.

13. U.S. Census of Housing: 1960 Series HC(3)-158
City Blocks of Cedar Rapids, Iowa
content: published table of data for city blocks, map of the city with city block lines.

B. Processing the Educational Data

Both Iowa and Florida possess what is termed here an "educational data bank center." These centers store and maintain vast amounts of data in machine readable form. In attempting to use these data banks for research purposes, problems arose due to the nature and activities of the centers themselves. "Processing" was necessary to make the data usable.

The School Units

Prelude to processing, it was essential to discover the data units, the entities to which information ascribed. A school unit is a single class of organizations or individuals described by educational data. A study of the Iowa data documents yielded for Iowa five school units:

County
District
School
Teacher
Student

The units District, School, Teacher, and Student are hierarchical units and posed no manipulative problems. In most cases, the district contained only one high school, and in these cases, the district and the school are equivalent units. In the case of the multi-high-school district, the school is considered as a school attendance unit (see Appendix B for more detailed definitions of district and school attendance units). The district and county are not geographically coterminous, but since the school is the focus of this research, and since educational data describing counties was of little or no value, the problem of dis-consonant units did not appear within the area of processing the educational data.

Problems with the Data Bank

An educational data bank center receives its funds from the state which it serves. Accordingly, the activities of the center are limited and oriented to very practical and essential applications--student transcripts, teacher record keeping, financial accounting, etc. These are not research applications.

From peripheral contact with the centers, a simplified model of their data collection and storing procedures was inferred; this is depicted in Figure 4-B-1. The source forms, for example, might be teacher personnel forms. Accompanying the actual forms would be perhaps a general information pamphlet meant to inform administrators and teachers of the purpose of the form; instructions for filling out the source forms might be contained in a separate manual or might appear on the source form itself. The completed source forms would be collected and coded on cards, for example. This original machine readable medium contained all the information that was coded from the forms.

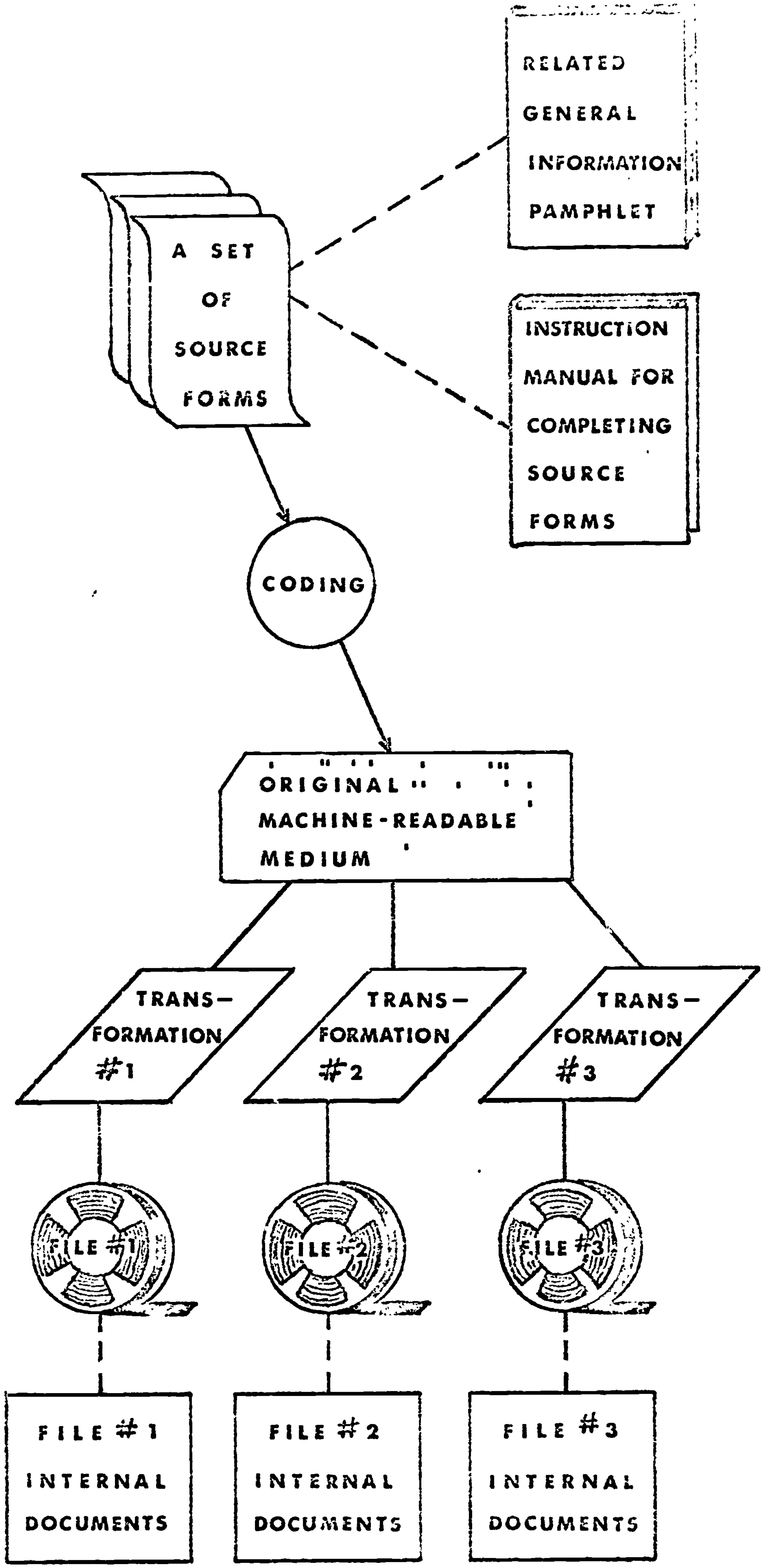


Figure 4-B-1
Model for Data Collection

Next, transformations were performed upon the original information: the order of the items of information might be slightly altered during the transformation; items of information might be deleted or recoded; items of information from two types of source forms might be merged together. Each transformation was designed for a very specific application, and each resulted in a second generation medium--a file. Further transformations might be performed on and between second or first generation media to produce third generation media, and so forth. What resulted was a diffusion of items of information and of transformed but equivalent items into many files.

Files were documented by data bank center personnel, but the documentation was minimal and designed for use mainly within the center. The first problem that arose then, was that no organized information was available to adequately answer the questions, "What information is stored? How is it coded?" In lieu of the transformations performed, a third question had to be made answerable: "Where is a particular item of information to be found?" The remainder of this section is devoted to a description of the procedures used to make these questions answerable--procedures used to make the data bank usable for research purposes.

The Iowa Format Book

The Iowa educational data bank center had thirty-eight files in its bank when a request for data information was made. Received from Iowa for each file was a set of internal documents (see Figure 4-B-1). Each file was documented by several pages including at least an identifying page and an information layout form. Sometimes included in the set was a sample of the source forms previously mentioned (see Appendix B, pp. 2-5 for an example of file documentation). The collection of the thirty-eight file documentations constitutes what hereafter is referred to as the "Iowa Format Book." The files of the Iowa Format Book were then analyzed for data content.

The Process of Content Analysis

Through the process of content analysis, data were made usable for research purposes. The analysis was concerned with "identifying" and recording the location of basic informational units contained in a collection of data. This basic informational unit of data is called an "item." For each item within each file, a Data Item File Sheet was completed (see Appendix B-Foldout for an example of a completed File Sheet). The following information was recorded on a File Sheet:

1. The school unit which the item described, i.e., student, teacher, school, district, or county. This unit applied to all items within the same file.

2. The abbreviated title of the item as it appeared on the information layout form.

3. A "key-word title" which summarized the character of the item.

   The key-word title included:

   a) the school unit,
   b) words giving the general nature of the item, e.g., salary or enrollment, and
   c) specific information about a) or b).

   The title would read as an ordinary nominal phrase. Synonymization within the title without loss of conciseness was desirable.

4. Any further description of the item.

5. The source used for the item's description (the sources were either the file documentation or one of the information manuals listed in Section A of this chapter).

6. The coding system for the item if applicable.

7. The source used in determining item six (#6)

8. Physical information--number and name of file.

Appendix B gives a detailed illustration of the procedure. Over 1000 items were processed in this manner. The File Sheets were then assigned identification numbers and certain parts of the item information were punched onto sets of computer cards (see Figure 4-B-2). The punched information for an item is called a "record", and its identification number, a RECID number.

Indexing the Item Information

Essentially, the Data Item File (collection of Data Item File Sheets) was a library of information pertaining to items of data. An index was necessary in order to locate information being sought. For instance, a standard book library will have its books identified with a Dewey Decimal System number, and will have a card catalogue with subject index, author index, and title index. Without these indexes, the library of books would be almost unusable. An unindexed Data Item File would be similarly unusuable.

A computer program was developed to automatically produce indexes from the punched version of the Data Item File. Examples of these indexes are given in Figures 4-B-3, 4-B-4, and 4-B-5. The produced indexes served a purpose analogous to that of the standard card catalogue. Immediate answers to the questions of data availability, coding, and locations of items were enabled through use of the indexed Data Item File.

1) KWIC index, Figure 4-B-3

The KWIC index is an alphabetical listing of all non-trivial words (key words) contained within the key word titles of the File Sheets. "KWIC" stands for Key Word In Context, and each line of the index has

```
IOWA0252KATEACHERS
IOWA0252TATEACHERS TOTAL SEMESTER HOURS OF COLLEGE CREDIT--TEACHERS PREPARATION
IOWA0252AA036-044
IOWA02520ATOTAL SEMESTER HOURS OF COLLEGE CREDIT EARNED AT UNDERGRAD AND GRAD LE
IOWA02520BVEL.  (QUARTER-HOURS OF CREDIT CONVERTED TO SEMESTER HOURS EQUIVALENT
IOWA02520CBY MULTIPLYING QUARTER HOURS BY 2/3.)  IA ED.DATA, P 5, ITEM 115.
```

```
IOWA0253KATEACHERS
IOWA0253TATEACHERS HIGHEST DEGREE--CREDENTIAL
IOWA0253AA036-044
IOWA02530ACODE INDICATING THE HIGHEST DEGREE HELD BY THE INDIVIDUAL.  IA ED DATA
IOWA02530B, P 5, ITEM 116.
```

```
IOWA0254KATEACHERS
IOWA0254TATEACHERS NUMBER OF DECLARED MAJORS--UNDERGRADUATE
IOWA0254AA036-044
```

Figure 4-B-2

Examples of punched versions of data item file sheets.

```
PECIAL COURSES FUND      DISTRICTS        SECRETARIES BOOK BALANCE FOR DISTRICTS S     IOWA2050
DITURES FOR THE BOARD OF EDUCATION AND    SECRETARY     DISTRICTS     DISTRICTS FEDE    IOWA2094
ADMINISTRATION--BOARD OF EDUCATION AND    SECRETARY     DISTRICTS     DISTRICTS GENE    IOWA2060
          CARD-NUMBER CODE FOR DATA OF    SECRETARYS ANNUAL REPORT     DISTRICTS        IOWA2007
                       YEAR OF DATA FOR   SECRETARYS ANNUAL REPORT     DISTRICTS        IOWA2006
L - SCHOOLS COURSES     SCHOOLS           SECTION ENROLLMENT BY SEX AND GRADE LEVE      IOWA0749
UDENTS - SCHOOLS COURSES     SCHOOLS      SECTION ENROLLMENT - TOTAL NUMBER OF FEM      IOWA0751
UDENTS - SCHOOLS COURSES     SCHOOLS      SECTION ENROLLMENT - TOTAL NUMBER OF MAL      IOWA0750
    SCHOOLS COURSES - LOCAL COURSE AND    SECTION NUMBER FOR SUBJECT     SCHOOLS        IOWA0737
 TEACHERS ASSIGNMENT--LOCAL COURSE AND    SECTION NUMBER FOR SUBJECT     TEACHERS       IOWA0276
TUDENTS                   PUPILS LOCAL    SECTION NUMBER IN COURSE DATA BLOCK     S     IOWA2166
       TYPE OF DISTRICT--HIGHEST GRADE    SECTION TAUGHT IN DISTRICT     DISTRICTS      IOWA2001
ES     SCHOOLS                            SECTION TOTAL ENROLLMENT - SCHOOLS COURS      IOWA0752
SFERS--RETURN OF PRINCIPAL OF INVESTED    SECURITIES AT COST     DISTRICTS     DISTR    IOWA2058
S FROM RETURN OF PRINCIPAL OF INVESTED    SECURITIES AT COST     DISTRICTS     DISTR    IOWA2044
ENERAL FUND CLOSING BALANCE-- INVESTED    SECURITIES     DISTRICTS       DISTRICTS G    IOWA0989
       DISTRICTS GENERAL FUND INVESTED    SECURITIES     DISTRICTS                      IOWA2055
ENERAL FUND OPENING BALANCE-- INVESTED    SECURITIES     DISTRICTS       DISTRICTS G    IOWA0908
GENERAL FUND TRANSFERS--MONEY PAID FOR    SECURITIES     DISTRICTS        DISTRICTS     IOWA2068
LHOUSE FUND CLOSING BALANCE-- INVESTED    SECURITIES     DISTRICTS     DISTRICTS SCH    IOWA0895
LHOUSE FUND OPENING BALANCE-- INVESTED    SECURITIES     DISTRICTS     DISTRICTS SCH    IOWA0862
STRICTS SCHOOLHOUSE FUND CASH PAID FOR    SECURITIES DURING THE YEAR     DISTRICTS      IOWA2047
STRICTS             DISTRICTS INVESTED    SECURITIES IN SPECIAL COURSES FUND    DI      IOWA2049
OLS                    TEACHERS SOCIAL    SECURITY NUMBER - IDENTIFICATION    SCHO      IOWA0739
ERS                    TEACHERS SOCIAL    SECURITY NUMBER--IDENTIFICATION    TEACH      IOWA0240
                TEACHERS PSEUDO SOCIAL    SECURITY NUMBER     TEACHERS                  IOWA0241
PARATION    TEACHERS    TEACHERS TOTAL    SEMESTER HOURS OF COLLEGE CREDIT--TEACHE      IOWA0252
UDENTS                    PUPILS FIRST    SEMESTER MARK IN COURSE DATA BLOCK     ST     IOWA2170
UDENTS                   PUPILS SECOND    SEMESTER MARK IN COURSE DATA BLOCK     ST     IOWA2171
SCHOOLS               SCHOOLS COURSES -   SEMESTERS IN WHICH COURSE IS OFFERED          IOWA0738
--COURSE DATA BLOCK     STUDENTS     THE  SEMESTERS THE PUPIL IS ENROLLED FOR A CO      IOWA2167
  SCHOOLS FEMALE GRADUATES EMPLOYED AS    SEMI-SKILLED LABOR--IN STATE     SCHOOLS      IOWA0498
    SCHOOLS MALE GRADUATES EMPLOYED AS    SEMI-SKILLED LABOR--IN STATE     SCHOOLS      IOWA0497
  SCHOOLS FEMALE GRADUATES EMPLOYED AS    SEMI-SKILLED LABOR--OUT OF STATE     SCHO     IOWA0500
    SCHOOLS MALE GRADUATES EMPLOYED AS    SEMI-SKILLED LABOR--OUT OF STATE     SCHO     IOWA0499
HOOLS LOWER HALF GRADUATES EMPLOYED AS    SEMI-SKILLED LABOR     SCHOOLS         SC     IOWA0503
   SCHOOLS TOTAL GRADUATES EMPLOYED AS    SEMI-SKILLED LABOR     SCHOOLS                IOWA0501
HOOLS UPPER HALF GRADUATES EMPLOYED AS    SEMI-SKILLED LABOR     SCHOOLS         SC     IOWA0502
     DISTRICTS GENERAL FUND RECEIPTS--    SEMI-ANNUAL APPORTIONMENT     DISTRICTS       IOWA0912
S NUMBER OF REMALE GRADUATES RECEIVING    SENIOR CERTIFICATE     SCHOOLS     SCHOOL     IOWA0414
OLS NUMBER OF MALE GRADUATES RECEIVING    SENIOR CERTIFICATE     SCHOOLS       SCHO     IOWA0413
LS TOTAL NUMBER OF GRADUATES RECEIVING    SENIOR CERTIFICATE     SCHOOLS      SCHOO     IOWA0415
```

Figure 4-B-3
Example of a KWIC Index of the Iowa Data Item File

```
STAFF/4,063-066                                                                    IOWA0245
    TEACHERS DATE OF BIRTH
    TEACHERS
STAFF/4,067-066                                                                    IOWA0246
    TEACHERS YEARS OF EXPERIENCE IN THIS DISTRICT
    TEACHERS
STAFF/4,069-070                                                                    IOWA2054
    TEACHERS TOTAL YEARS OF EXPERIENCE
    TEACHERS
STAFF/4,071-073                                                                    IOWA0247
    TEACHERS CONTRACT PERIOD LENGTH--TEACHERS EMPLOYMENT
    TEACHERS
STAFF/4,074-076                                                                    IOWA0248
    TEACHERS PERCENTAGE OF FULL TIME EQUIVALENCY--TEACHERS EMPLOYMENT
    TEACHERS
STAFF/4,077-081                                                                    IOWA0249
    TEACHERS SALARY--EARNINGS--TEACHERS EMPLOYMENT
    TEACHERS
STAFF/4,082-082                                                                    IOWA0250
    TEACHERS EXTRA COMPENSATION--EARNINGS--TEACHERS EMPLOYMENT
    TEACHERS
STAFF/4,083-083                                                                    IOWA0251
    TEACHERS OCCUPATION OF PREVIOUS YEAR--TEACHERS HISTORY
    TEACHERS
STAFF/4,084-086                                                                    IOWA0252
    TEACHERS TOTAL SEMESTER HOURS OF COLLEGE CREDIT--TEACHERS PREPARATION
    TEACHERS
STAFF/4,085-086                                                                    IOWA0266
    TEACHERS CLASS OF CERTIFICATE--LENGTH OF TIME OF CERTIFICATION
    TEACHERS
STAFF/4,087-087                                                                    IOWA0253
    TEACHERS HIGHEST DEGREE--CREDENTIAL
    TEACHERS
STAFF/4,088-088                                                                    IOWA0254
    TEACHERS NUMBER OF DECLARED MAJORS--UNDERGRADUATE
    TEACHERS
```

Figure 4-B-4

Example of an Author Index of the Iowa Data Item File

```
IOWA0248 *** STAFF/4,074-076                                                    IOWA0248
    TEACHERS PERCENTAGE OF FULL TIME EQUIVALENCEY--TEACHERS EMPLOYMENT
    TEACHERS
    FULL-TIME EQUIVALENCY OF PART-TIME PERSONNEL EXPRESSED AS A PERCENTAGE. IA ED DATA, P 5, ITEM 111.

IOWA0249 *** STAFF/4,077-081                                                    IOWA0249
    TEACHERS SALARY--EARNINGS--TEACHERS EMPLOYMENT
    TEACHERS
    TOTAL PROFESSIONAL CONTRACT SALARY, INCLUDING PAY FOR EXTRA DUTIES AND OTHER COMPENSATION WHICH MAY BE
        IN ADDITION TO BASE SALARY.  IA ED DATA, P 5, ITEM 112.

IOWA0250 *** STAFF/4,082-082                                                    IOWA0250
    TEACHERS EXTRA COMPENSATION--EARNINGS--TEACHERS EMPLOYMENT
    TEACHERS
    CODE INDICATING REASON OR REASONS FOR INDIVIDUALS SALARY EXCEEDING THE NORMAL STEP ON THE SALARY SCHEDULE
        IA ED DATA, P 5, ITEM 113.

IOWA0251 *** STAFF/4,083-083                                                    IOWA0251
    TEACHERS OCCUPATION OF PREVIOUS YEAR--TEACHER HISTORY
    TEACHERS
    CODE INDICATING INDIVIDUALS OCCUPATION DURING PREVIOUS SCHOOL YEAR.  IA ED DATA, P 5, ITEM 114

IOWA0252 *** STAFF/4,084-086                                                    IOWA0252
    TEACHERS TOTAL SEMESTER HOURS OF COLLEGE CREDIT--TEACHERS PREPARATION
    TEACHERS
    TOTAL SEMESTER HOURS OF COLLEGE CREDIT EARNED AT UNDERGRAD AND GRAD LEVEL. (QUARTER-HOURS OF CREDIT
        CONVERTED TO SEMESTER HOURS EQUIVALENT BY MULTIPLYING QUARTER HOURS BY 2/3.) IA ED DATA, P 5, ITEM 115.

IOWA0253 *** STAFF/4,087-087                                                    IOWA0253
    TEACHERS HIGHEST DEGREE--CREDENTIAL
    TEACHERS
    CODE INDICATING THE HIGHEST DEGREE HELD BY THE INDIVIDUAL.  IA ED DATA, P 5, ITEM 116.
```

Figure 4-B-5

Example of a RECID Index of the Iowa Data Item File the key word centered within a portion of its original context. One line appears for each occurrence of a key word, and the RECID (record identification number) of the titled item appears to the right of the line. The full entry can be found by locating this RECID number in the RECID index, Figure 4-B-5.

A KWIC index lends itself ideally to the clustering of available data. The researcher may have a theoretical concept in mind and ideas as to what might be measures of that concept. He can instantly find available and relevant items of data by referencing words in the KWIC index that are descriptive of the measure. He may be motivated to new ideas for measuring by a directed search through the index. For example, if information about high school seniors is desired, one could find all the references with the word "senior" in the title by simply finding the word. After reading the title segments, one notices that "graduates" is also a good place to look for this information.

2) Author index, Table 4-B-4

This index gives a list of all the items arranged in order of file and of physical location within the file. The library card catalogue author index is analogous to this index: items ordered by location within a file are analogous to books ordered by title within an author.

The first line of each three-line entry is the file name followed by a slash followed by the item's physical location. The RECID again appears on the right. The second line is the item title, and the third, the school unit.

A reconstruction of the Iowa Format Book including references to complete item documentation was obtained through production of the Author index.

3) RECID index, Table 4-B-5

This index is ordered according to its record identification number, and each entry contains all the information punched for the item. Of course, the actual File Sheet may be located by use of the RECID number, and full documentation including coding can be found there.

Selecting the Data

The produced indexes made answerable the questions about availability of data, coding, and location. A study of the indexes revealed that five files were exhaustive, i.e., five of the thirty-eight files contained practically all the information in the bank. These files were requested and were received from the Iowa Information Center. Four of the five files were contained on reels of magnetic tape; the fifth, on cards.

Finalizing the Item File

Since only a sub-set of the entire bank of items was acquired, the Data Item File had to be modified accordingly and then reindexed. From the original set of punched records of item information, and File Sheets, a selection pertaining only to the files received was made. Certain additions had to be made since the data received were of a different school year than were the content analyzed file documents. Other modifications to provide for a precise physical location of items had to be included for later programming purposes.

The finalized selection of punched card records was then run through the indexing program, and final indexes were produced.

C. Processing the Census Data

Resolutions to the problem of data and the problem of data units were brought about through the processing of census data. The first problem, that of data, was resolved by applying similar procedures to census data as were applied in processing the educational data (section B of this chapter). A comparatively small effort was required in this area of processing, however, since census data are highly organized and well documented. The second problem, that of data units, was concerned with bringing data available for census units into consonance with data available for school units. As was previously mentioned (Chapter 1, Section C), school district boundaries do not follow any census unit boundaries. Thus, transformations had to be performed upon census unit data to enable compatibility with school unit data. The "mapping" of the Iowa school districts provided the basis for this transformation.

Census Units

A complete list of census definitions is given in Appendix A, but those listed below are sufficient for an understanding of this section.

Counties

> The primary divisions of the States except for Louisiana where the divisions are called parishes and in Alaska where the data are collected for election districts.

Minor Civil Divisions

> The primary political divisions into which counties are divided. Where more than one type of primary division exists in a county, the bureau uses the more stable division so that comparable data are available from census to census (school, taxation, election units, etc. are not considered stable).

Standard Metropolitan Statistical Areas (SMSA)

> A unit whose entire population is in and around the city whose activities form an integrated social and economic system. Except in New England, an SMSA is a county or group of contiguous counties which contains at least one

city of 50,000 inhabitants or more or "twin cities" with a combined population of at least 50,000. In addition to the counties containing such a city or cities, contiguous counties are included in an SMSA, if, according to certain criteria, they are essentially metropolitan in character and are socially and economically integrated with the central city. Criteria are listed on page viii of PC(1).

Census Tract

Census tracts are small areas into which large cities and adjacent areas have been divided for statistical purposes. Tract boundaries were established cooperatively by a local committee and the Bureau of the Census, and were generally designed to be relatively uniform with respect to population characteristics, economic status, and living conditions. The average tract has about 4,000 residents. Tract boundaries are established with the intention of being maintained over a long time so that comparisons may be made from census to census.

Place

Concentrations of population, regardless of the existence of legally prescribed limits, powers, or functions. Most of the places listed are incorporated as cities, towns, villages, or boroughs, however. The larger unincorporated place outside the urbanized area was delineated and those with a population of 1,000 or more are presented in the same manner as incorporated places of equal size.

Enumeration District

A small area assigned to an enumerator which must be canvassed and reported separately. In most cases an ED contains approximately 250 housing units. The boundaries for the ED's for the most part follow such features as roads, streets, railroads, streams, and other clearly defined lines which may be easily identified by census enumerators in the field and often do not conform to the boundaries of political units.

Availability and Selection of Census Data

Census information is divided in three ways: major subject fields, geographic areas, and retail trade areas (see Appendix A, p. 5). Of the eight major subject fields, population and housing data were considered to contain the best selection of items which could characterize a community.

Booklets describing data availability were readily obtained along with other census publications. These are listed in Table 4-A-2. Data obtainable in the form of magnetic computer tape reels are available at small census unit levels--enumeration districts, census tracts, and minor civil divisions.

The data are of one of two types according to the method of collection--count data and sample data. The count data are based on a 100% sample while the sample data are based on a sample of approximately 25%; calculations are conducted by the Census Bureau to bring sample figures up to an estimate of total population figures. No sample data were tabulated for enumeration districts because of sampling variability at the level of this smallest of census units. A list of the subject categories of data items available for count data and for sample data is given below:

Subjects Available for Count Data
(available for enumeration district, census tract, minor civil division)

1. Age
2. Household Relationship
3. Marital Status
4. Race

Subjects Available for Sample Data
(available for census tract, minor civil division)

1. Age
2. Class of Worker
3. Country of Origin
4. Employment Status
5. Income of Families and Unrelated Individuals
6. Industry
7. Marital Status
8. Married Couples, Families, and Unrelated Individuals
9. Means of Transportation
10. Nativity and Parentage
11. Occupation
12. Place of Work
13. Residence in 1955
14. Residence: Urban-Rural, Farm-Non-farm
15. School Enrollment
16. Years of School Completed

So, although the smaller unit was desirable, both sets of data could not be obtained for enumeration districts. As will be seen later in the mapping of Iowa school districts, a decision with large ramifications was at hand in selecting the data. Even though non-dollar costs were of main interest in the research, actual dollar costs had to be considered in the course of research activities. The considerations of time and cost made clear the decision as to which units of availability the data should be chosen. The MCD (minor civil division) and census tract units were the only practical choice. This choice also meant a wider variety of community data--the entire scope presented above.

Indexing the Census Data

Along with the magnetic tape reels of actual data, two technical memoranda explaining the record layout and content of the sample and count data were received. From these two documents, a census item file was constructed by completing a census data item file sheet for each item contained on the tapes; the completed file sheet contained the title of the item and its exact physical location on the computer tape. This information was punched on sets of computer cards and indexes of the census data information were produced automatically.

The procedures presented so far in this section have closely paralleled those of the previous section, processing the educational data. And at this point, with the index refined, a means of instantaneous access to information concerning both educational and community data existed.

Disconsonant Units

As was stated earlier, the school unit is not geographically coterminous with census units. The "disconsonance" is best discussed by classifying the school district according to multiple or single high school district, and location of the district inside or outside an SMSA. Table 4-C-1 lists the four cases. The classification is a useful aid to discussion, but is imperfect since school districts may have area in two or more counties and, therefore, be neither entirely inside or outside an SMSA.

Case I. Single High School, Non-SMSA

Geographically, high school attendance areas for districts with one high school unit are equivalent to the district itself. Therefore, the major educational data unit in this case is the district. The major census unit is the township, sometimes referred to as the MCD (minor civil division). Figure 4-C-1 represents a section of central Iowa and exemplifies Case I. Note that the school district boundaries do not follow the township lines. The majority of the disconsonance was of this type.

Nothing was contained in either the census data or the educational data to indicate the correspondences between MCD's and school districts. Maps displaying both units had to be studied in order to discover proportional correspondences.

Case II. Multiple High School District in SMSA

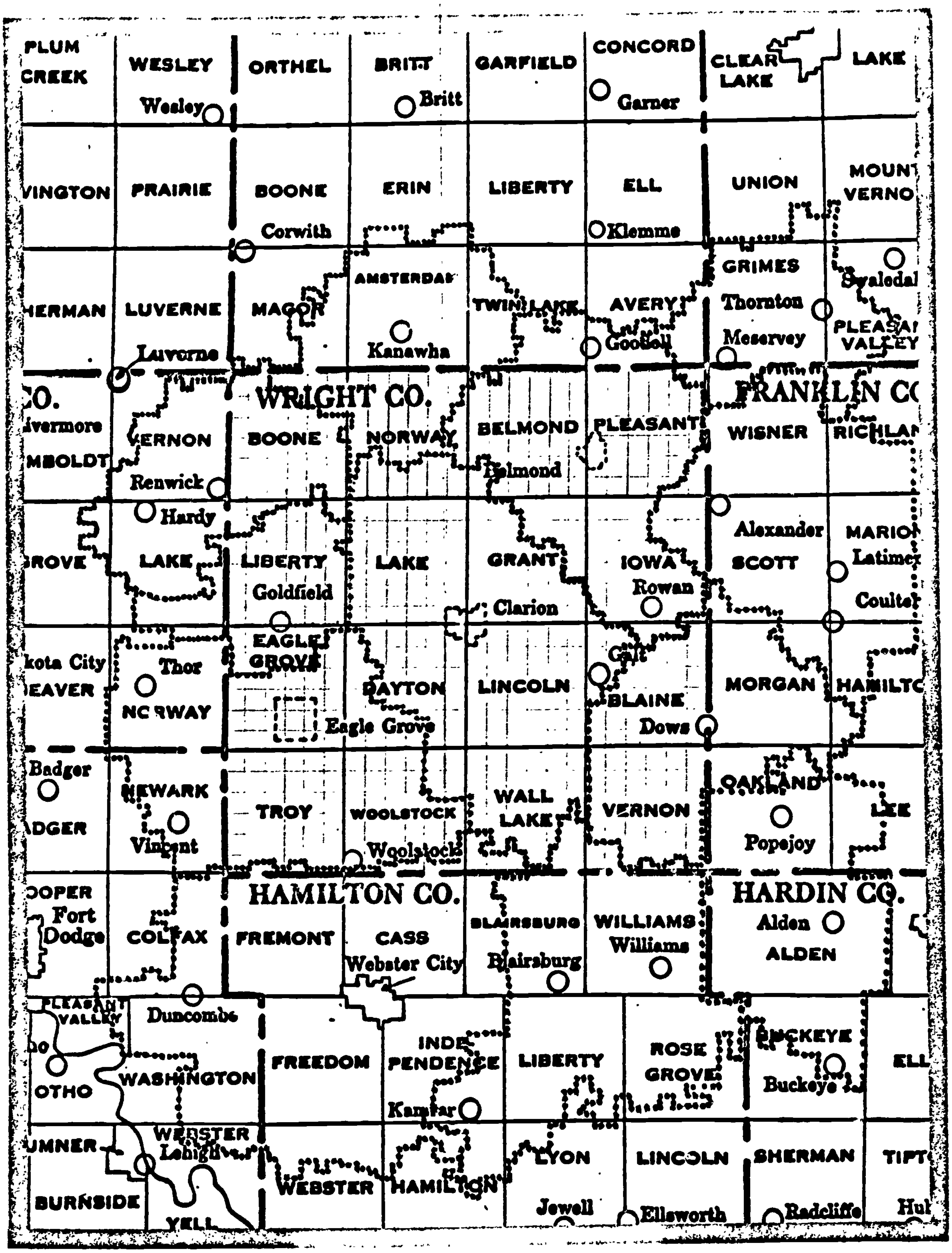
The state of Iowa possesses five SMSA cities, i.e., five cities within tracted araas. The school district in each of these cities is a multiple high school district. In this case the major educational data unit was the high school attendance area, and the census data unit was the census tract. Figure 4-C-2 is a census tract map of Cedar Rapids, Iowa and exemplifies Case II. In most instances, census tracts were completely contained within one high school attendance area, but in outlying districts where tracts became larger, school attendance areas cut across tracts.

Again, maps with attendance areas drawn on them had to be acquired and studied before proportional correspondences could be discovered.

TABLE 4-C-1

CASES OF DISCONSONANT UNITS

	Single High School District	Multiple High School District
District Outside an SMSA	Case I. School Unit = district Census Unit = MCD Figure 4-C-3	Case IV. School Unit = high school attendance area Census Unit = MCD
District Inside an SMSA	Case III. School Unit = district Census Unit = census tract	Case II. School Unit = high school attendance area Census Unit = census tract Figure 4-C-4



-LEGEND-

County line:

Township line:

Township section:

School district boundary:

Figure 4-C-1

Geographical Areas of Wright County, Iowa, with School Districts Superimposed.

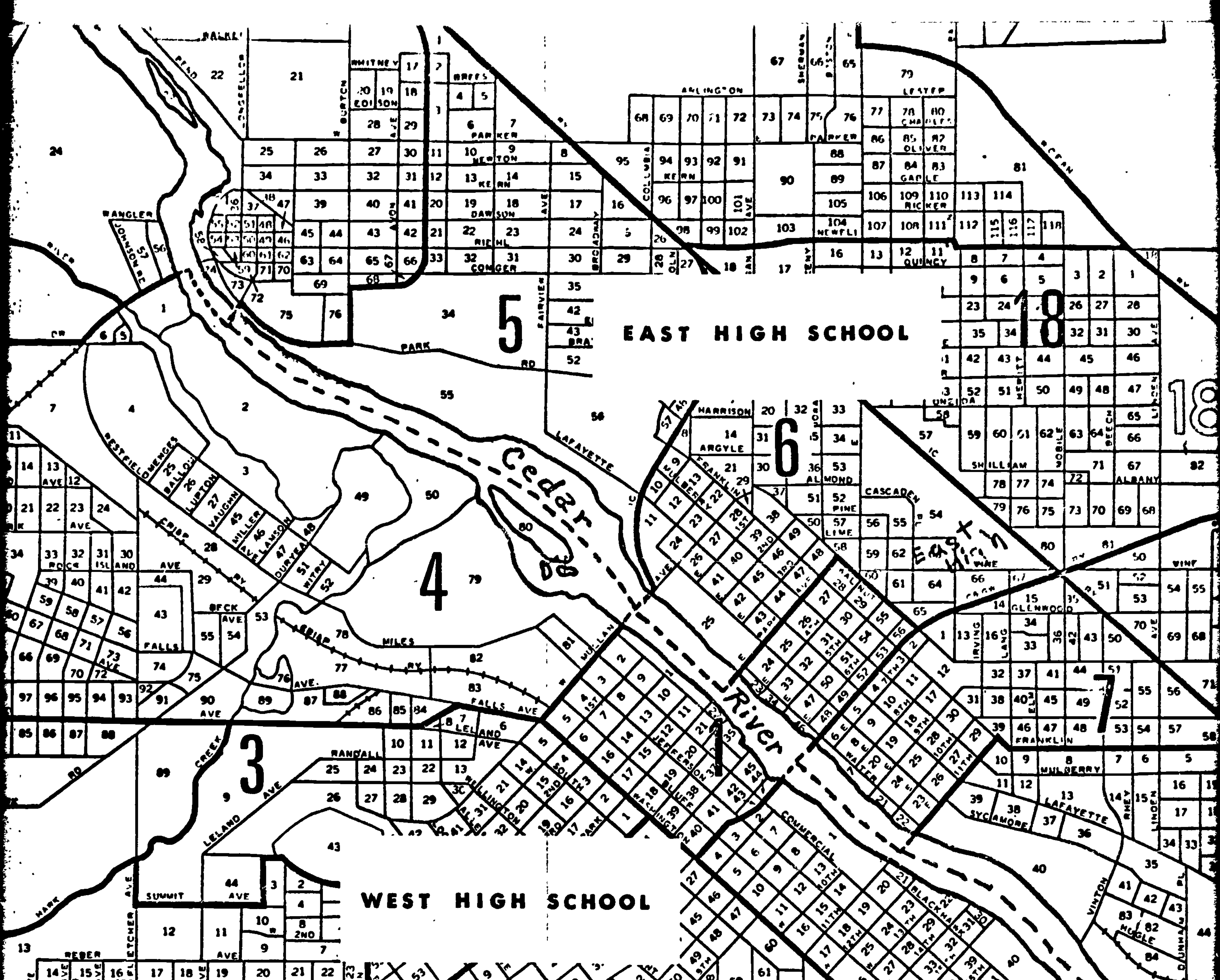


Figure 4-C-2

Disconsonance of Units, Case II

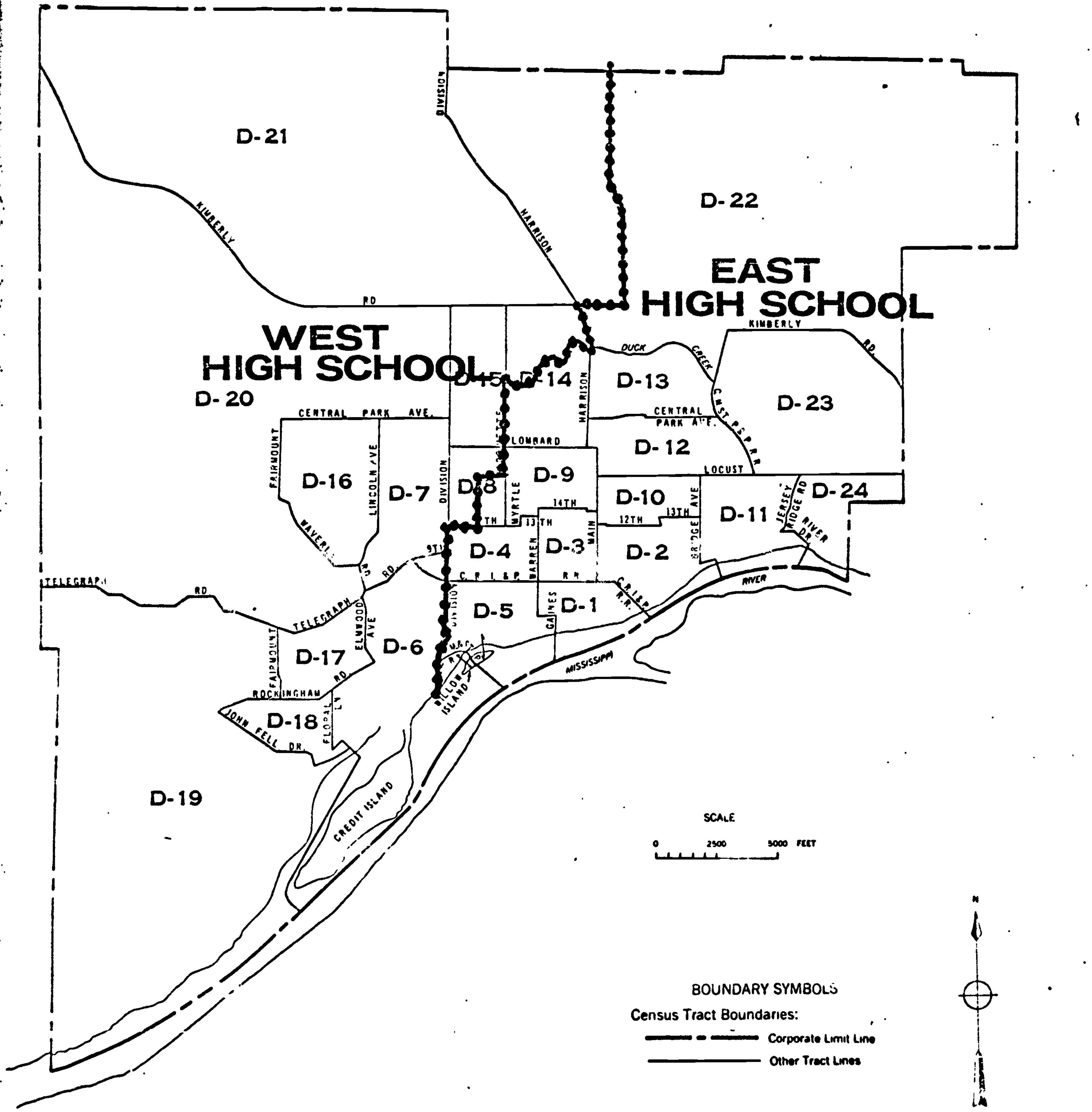


Figure 4-C-3

Disconsonance of High School Attendance Area and Census Tract

Case III. Single High School District in SMSA

The districts outside the SMSA city but within the SMSA county were all single high school districts. The school unit was therefore the district, and the census unit, the census tract. The outlying tracts were as large or larger then MCD's, and Figure 4-C-3 gives an idea of the relative sizes of the disconsonant areas.

Case IV. Multiple High School District, Non-SMSA

There was only one occurrence of a multiple high school district in a non-tracted area. The disconsonance was between high school attendance area and MCD. Figure 4-C-1 is again useful to understanding the type of disconsonance if the shaded area is considered to be a single district with two high schools.

A Mapping Solution to the Problem of Disconsonant Units

How can census data be accurately allocated to schools or school districts if the proportional correspondences are not recorded except in pictorial form? After much deliberation, it was decided that the only practicable approach to the problem was to establish which census tracts, towns, cities, townships or parts of them are included within the various school district boundaries. The cities, towns, and smaller census tracts posed no particular problem since they were located completely within the area of the school unit, but the townships and outlying census tracts because of their size had to be parceled to two, three or four different school districts. The most obvious basis for splitting up the townships was that of proportional area contained in the various school districts. But was this basis justifiable?

Professor Randall Sale of the University of Wisconsin Geography Department was asked to do some research on this question. Professor Sale selected three townships in Iowa with widely differing terrain and had dwelling units plotted on the maps of the townships to test an assumption that population is a continuous and uniformly distributed variable. After plotting the areas and studying the relationship between the terrain and the population distribution, evidence indicated that defining population on the basis of area was a fair assumption. In other words, one can draw lines through Iowa townships in almost any way, and the population will be divided in much the same way as the area. Keep in mind, however, that this method is used with townships in rural areas or with census tracts in highly urbanized areas, if partitioning is necessary there.

The Mapping Task

The U.S. Census Geographic Code Scheme Booklet (Table 4-A-3,#4) contained the names and identification code numbers of all the geographic census areas in Iowa, e.g., counties, townships, towns, parts of towns in a township, etc. With the aid of a computer, forms depicted in Figure 4-C-4 were produced. The next step was to travel to Des Moines with the forms, study a farrago of school district maps prepared by county superintendents, and record the proportional assignments of area (the italic of Figure 4-C-4). Census tracts for Iowa's five SMSA's were assigned similarly to high school attendance areas.

More than four hundred townships were apportioned in approximately one hundred man-hours of work. Local people were trained on the job in Des Moines and were supervised by a research staff member. The accuracy of certain of the areal apportionments was questionable, however, because of lack of definition in some of the maps submitted by the county superintendents.

Transformation of Census Data

Basically, two related lists had been compiled: the first was a list of all the geographic areas in Iowa--areas which had been assigned to school districts; the second was a list of actual census data units--units for which data were stored on magnetic computer tape.

Figure 4-C-4 is the list of geographical areas and proportional assignments for Wright County, Iowa. These areas are displayed in Figure 4-C-5 with school districts superimposed.

The areas preceded by an asterisk in Figure 4-C-4 are also actual census data units--units for which data were stored on tape. The units were of three kinds:

If the township contained cities or parts of cities of population 2,500 or greater, the units for the township were

1) Cities or parts of cities of population 2,500 or greater.

2) The remainder of the township, i.e., the township minus 1)
   If the township had no large cities, the unit for that township was
3) The entire township.

With these two lists, enough information was present to automatically determine the census unit composition of each school unit. For example, from looking at Figure 4-C-4 one sees that school district 594 is composed of: (see also Figure 4-C-1)

97% of the remainder of Belmond Township
+100% of Belmond City
+ 22% of Blaine Township
+ 30% of the remainder of Grant Township

A computer was programmed to perform a more complicated set of operations than those indicated above, and the problem of disconsonant units was finally resolved. Using the results of mapping as a basis for the transformation, census data were allocated to the proper school districts. A final computer tape resulted which contained the proper community data to describe each district.

Identification Code Numbers of Area	Area Name	District Numbers	% of Area in District	District Numbers	% of Area in District	District Numbers	% of Area in District
42099	WRIGHT COUNTY						
*42099001	BELMOND TWP	1206	3%	594	97%		
*420990010365406O	BELMOND CITY (PART)	594	100%				
*42099002	BLAINE TWP	1206	19%	594	22%	1854	59%
420990021205402l	DOWS TOWN (PART)	1854	100%				
4209900215604001	GALT TOWN	1854	6%	1206	94%		
*42099003 7 1	BOONE TWP	729	100%				
*42099004	DAYTON TWP	1206	60%	1944	40%		
*4209900407704060	CLARION CITY (PART)	1206	100%				
*42099005	EAGLE GROVE TWP	1944	70%	2529	30%		
*4209900512654060	EAGLE GROVE CITY	1944	100%				
*4209900516504021	GOLDFIELD TOWN (PART)	2529	100%				
*42099006	GRANT TWP	1206	70%	594	30%		
*4209900607704060	CLARION CITY (PART)	1206	100%				
*42099007	IOWA TWP	1206	2%	916	20%	594	78%
4209900737504011	ROWAN TOWN	594	100%				
*42099008	LAKE TWP	1206	100%				
*4209900807704060	CLARION CITY (PART)	1206	100%				
*42099009	LIBERTY TWP	2529	80%	732	20%		
4209900916504021	GOLDFIELD TOWN (PART)	2529	100%				
*42099010	LINCOLN TWP	1206	100%				
*4209901007704060	CLARION CITY (PART)	1206	100%				

Figure 4-C-A

Proportional Assignments of Census Areas To School Districts

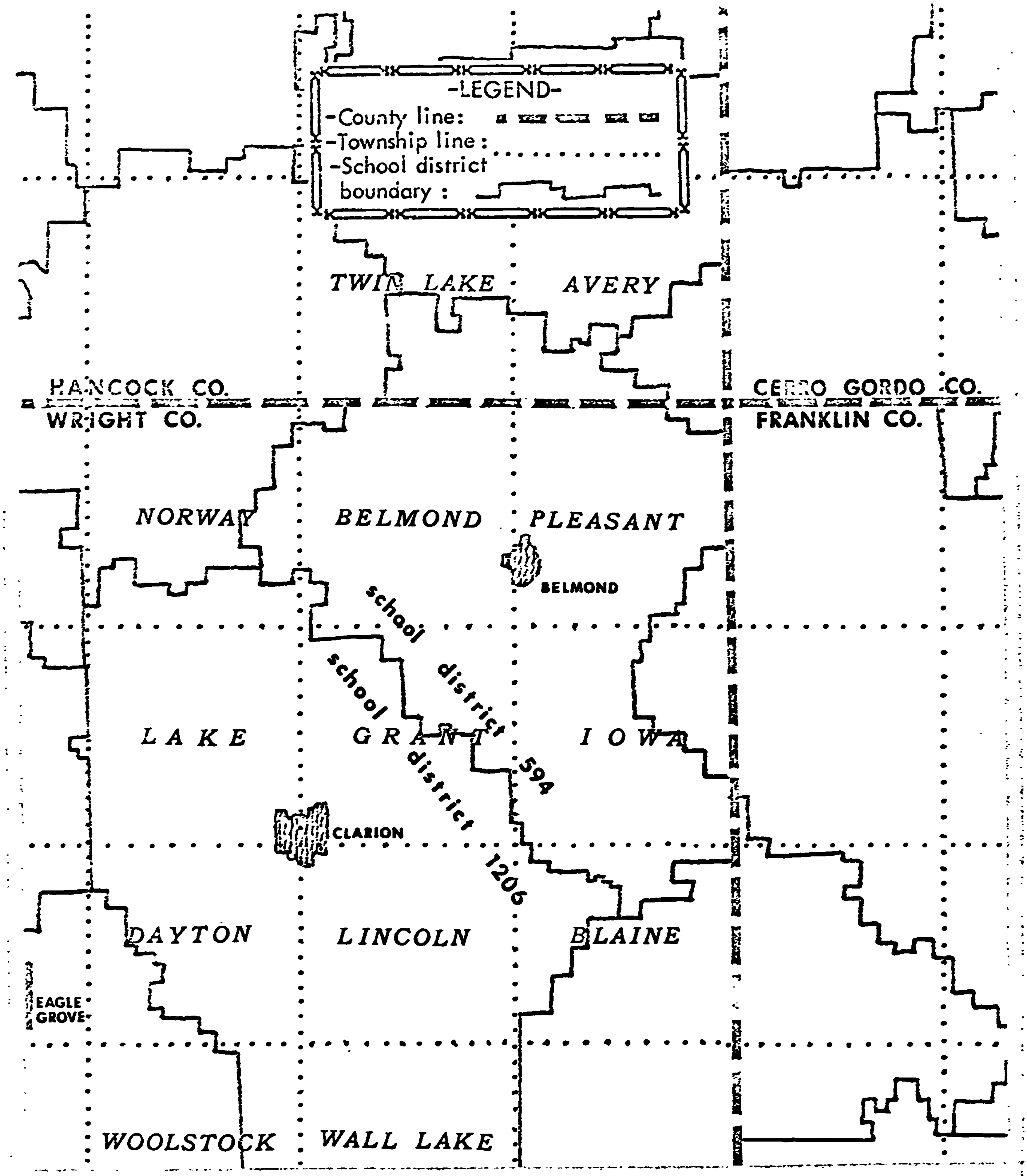


Figure 4-C-5

Disconsonance of District and Township

CHAPTER 5

SPECIFIC ANALYTICAL AND THEORETICAL TECHNIQUE

In this chapter the particular theoretic constructs and analytic operations are detailed. In Section A, a theory of the relationship between high school size and student output is presented; in this theory, clusters (characteristics, processes, qualities, etc.) are defined which together manifest a conceptualization of the substantive problem. In Section B, specific notes are made on how ideally these clusters might be measured and on the degree to which such ideal measurement can be realized with the presently available data. Then in Section C is presented a partitioning and grouping of the available data, as indicators, into the defined clusters. Finally, in Section D, the particular empirical operations then applied to the partitioned data are specified. The results and interpretations of the results of the operations are presented in Chapter 6.

A. A Theory for Conceptualizing Student/School Interaction

The theory presented in this section is intended to meet two general goals: it was necessary to establish a theoretical framework or understanding of the processes and factors at work in the relationship between high school size and non-dollar costs, first, in order to evaluate the adequacy of the data bank contents and, second, in order to perform actual data analysis. The theory consists of a series of defined processes and characteristics which are considered to be aspects of the substantive problem. The characteristics and processes are defined in such ways that a particular school may be considered to include "high" or "low" amounts of them: the characteristics and processes ultimately refer to continuous scales. In addition, the theory contains general causal chains connecting the characteristics and processes. In terms of the first goal, evaluating the adequacy of the data bank contents, the characteristics and processes define areas which should ideally be measurable from the data banks. In terms of the second goal, setting up data analysis, the characteristics and processes are used as categories around which actual data are clustered. Because of this second usage, the generic term "cluster" is adopted for referring to the defined characteristics and processes of the theory.

Because the theory is intended to encompass the entire substantive problem, it draws upon many scientific fields of study--for example, social psychology, social area analysis, educational administration, and mental testing. The theory is proposed as an <u>ad hoc</u> tool for a practical research problem, and the investigators realize that it is necessarily imperfect and that it is not unique. Two general principles guided its development. First, there was concern at crucial points in the theory for describing clusters which related to primary causes of student growth and for avoiding the use of constructs standardly defined for the purposes of administrative simplicity and utility. For example, the standard financial classifications within which schools operate are not considered important *per se* in describing the qualities of a school, because they affect the students only as secondary causes. The second principle states that sufficient descriptive complexity must be made of the aspects of the problem. In certain less crucial parts of the theory, in fact, the major concern was for sufficient complexity, the hope being that more meaningful (primary)clusters were transformationally related to the defined clusters. (This is explained in more detail later.)

Four general areas or superclusters are defined within the theory: Community Input, School Mediation, Student Output, and School Manipulation. The presentation of the theory is made in five parts. First, the superclusters are initially defined and their general causal interrelationships are explained. Then the four superclusters are presented one at a time. After the last page of this section there is a foldout page which contains a large diagram, Figure 5-A-1, of the clusters and cluster interrelationships of the theory. It is suggested that the diagram be folded out while the section is being read.

Overview of the Theory

The unit ultimately under study is that of an individual high school student, and the process ultimately under study is that of a student's growth. A student grows in an environment which includes his family, his peers, his church, his town, and his school. For the purpose of establishing a theory of a student's growth, four complexes of factors are identified:

(a) School Mediation, the processes which occur within the high school;

(b) School Manipulation, the secondary factors which affect School Mediation;

(c) Community Input, the non-school influences in the environment;

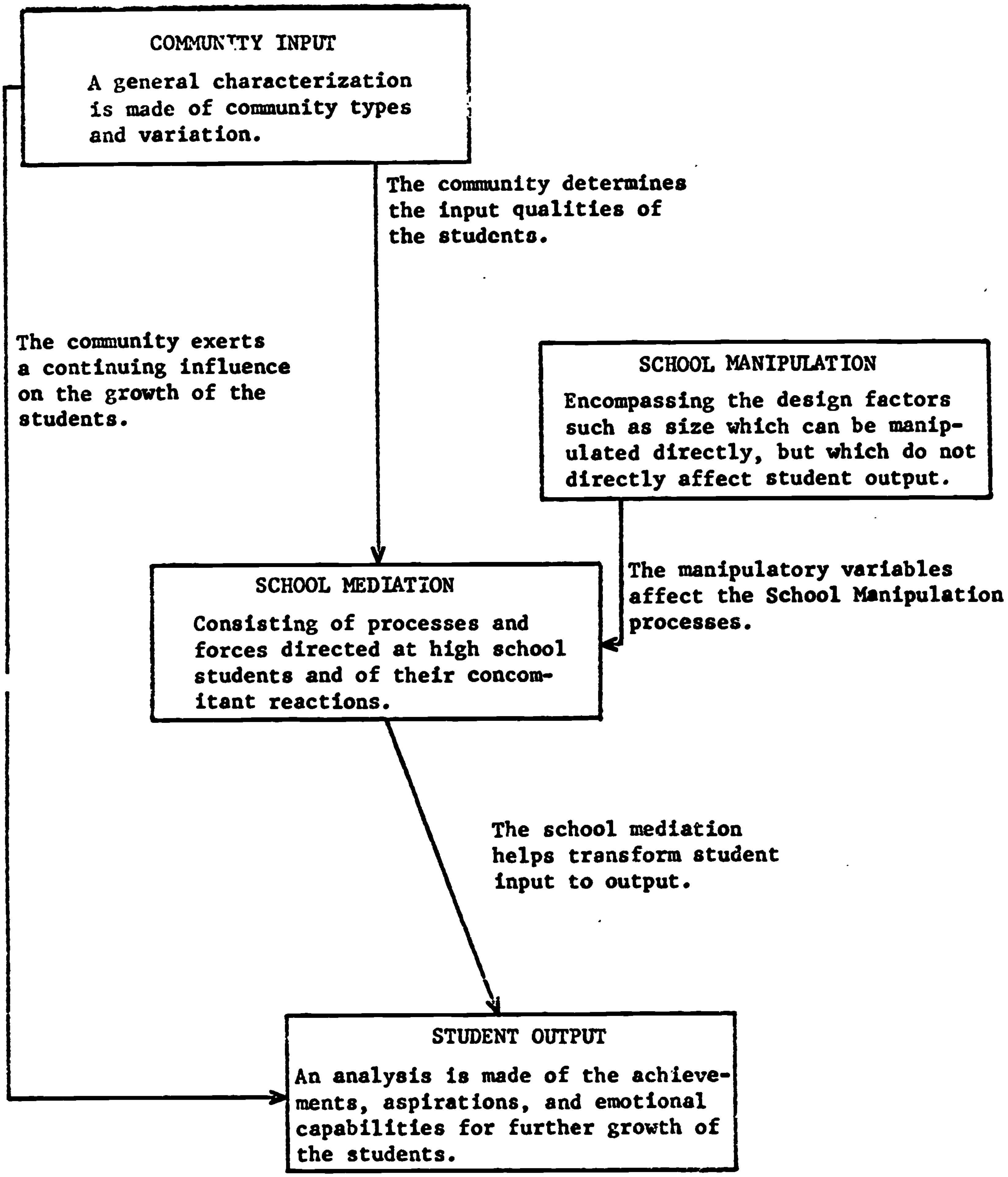
(d) Student Output, the effect on the student.

Their general interrelationships are diagrammed in Table 5-A-1. These four superclusters correspond to the four main boxes in the foldout. The arrows connecting the boxes indicate the interrelationships of the superclusters.

Community Input includes family and home influences and the effect of pre-high school growth. Therefore, the characteristics of the community represent input characteristics of the student body in the high school.

TABLE 5-A-1

THE FOUR MAIN SUPERCLUSTERS AND THE THEORETIC INTERRELATIONSHIPS



For example, the quality of the elementary schools in a community influences the initial scholastic abilities of high school students. But also, the community exerts a continuing and immediate influence on the students while they are in high school. For example, the existence of a library in a community may influence the academic growth of students. There are certainly factors in both the initial and the continuing influence of community which could be considered primary in their effect on students. Family heritage, elementary education, and community social process could conceivably be combined and categorized according to factors which directly affect the nature and growth of a student. However, for reasons explained later, the Community Input factors are defined rather independently of student relatedness; communities are characterized according to demographic variation.

School Mediation is a particular complex of influences within a more general set of community influences, but the present development of the theory of School Mediation has been isolated and is more exacting. School Mediation consists of the forces exerted on, or at least directed at, the student and of the concomitant reactions of the student. A complete understanding of the relationships between high school size and student growth depends on an understanding of the exact nature and strength of the forces and reactions of School Mediation as they differentially relate to school size. The clusters which are developed in the theory for School Mediation are intended directly to represent processes which primarily cause student growth.

The Student Output of the community and high school are the qualities of the students that emerge from them. The concern is for the characteristics of the individual student--his achievements, aspirations, and his emotional stability--for further growth. Because this study relates primarily to the mediation provided by the school, the outputs associated formally or informally with the school are emphasized. For example, ability to perform a job consonant with the opportunities of the community might be an output. Whether or not a student obtains a job might involve factors outside the scope of the theory.

The general causal chain of the theory has now been outlined: Community Input determines the qualities of the students upon entrance to school and also exerts a continuing influence; School Mediation is a particular influence and is to be examined in detail; Student Output is the dependent result. The fourth is a complex of potential factors which are grouped under School Manipulation. In determining how to design, build, or run a high school there are certain variables which may be manipulated by school administrators or boards of education. Color of the walls, student/teacher ratio, and number of rooms are examples. School size, in terms of the enrollment, is the School Manipulation variable upon which the present research is focused. It is to be emphasized, however, that School Manipulation--and, in particular, school size--are not considered part of the primary causal chain. Essentially, school size does not affect Student Output; rather it affects School Mediation which in turn affects Student Output. It has been necessary to establish the primary causal chain of Input, Mediation, Output in order to determine the ways in which size indirectly affects Student Output; in order to investigate the "non-dollar" costs of size.

In the rest of this section, the clusters within each of the four superclusters will be defined. One further general note should be made, however, on the problem of choice of unit. The theory ultimately is to be applied to schools; that is, the clusters to be defined are to be measurable for schools, each school having a single value. Yet as mentioned, the ultimate unit of the theory is the individual student, and the concern is for forces exerted on him and for his reactions and qualities. The unit of application, schools, seemingly has as its "value" for a characteristic a distribution of numbers across students rather than a single number for a school. When this is of concern, as it is for School Mediation and Student Output, two clusters are defined for a given initial cluster. One describes the mean-level value for the individuals in a school; one describes the variation of the value across students. It is assumed that a sufficient description is thus made of the distribution of values.

Community Characteristics as Input

As mentioned, the Community Input clusters represent two parts of the causal chain. First, they represent the characteristics of the student population upon input to high school, for the earlier growth of input students is carried out within and therefore determined in part by the community. The wholistic community clusters indicate the distribution of individualistic student characteristics. For example, a high percentage of farmers in a community indicates a high number of students aspiring to be farmers. The community clusters are assumed to represent indirectly the abilities, aspirations, and emotional outlook of the students as they enter high school.* Second, the Community Input clusters represent the continuing influences of the community, including the family, while the students are in high school. (School Mediation includes continuing influences of the community which have been segregated for detailed study.) The community clusters are assumed to indicate such non-school factors as family pressures and community activities which help mold the output characteristics of a student. For example, students may be distracted from their high school studies by family economic problems, their choice of study may be influenced by particular occupational opportunities of local industries, or community social pressure might stimulate academic achievement. These clusters are given below:

A. Population Dynamics	A1. Growth
	A2. Youth
	A3. Density
B. Population Characteristics	B1. Rural/urban
	B2. Socioeconomic Level
	B3. Socioeconomic Variation
C. Socialization Characteristics	C1. Social Activity/Involvement
	C2. Social Mobility
	C3. Poverty Culture

*Compare the clusters defined for Student Output.

D. Economic Structure

D1. Wealth
D2. Industry
D3. Trade
D4. Farming

E. Education Resources

E1. Archive Usage
E2. Educational Involvement
E3. Educational Attainment

Community Input clusters could conceivably be defined according to factors which directly correspond to characteristics of students and continuing influences of the community.* Such clusters could be aligned according to causes of characteristics and influences; parental encouragement toward higher education, parental intelligence, and a student's activity in community social projects might then be clusters. Or the clusters could be aligned according to the effects of the characteristics and influences; then cognitive ability, emotional stability, and aggressiveness might be clusters. In the present theory, however, the Community Input clusters are defined according to "social area analysis" community factors; the clusters are not designed directly to represent particular characteristics of students or particular influences on them, but rather to characterize the general demographic and social qualities of communities. This approach was taken because a partitioning according to primary causes would have required too great a degree of theoretical understanding of the growth processes", and a partitioning according to primary effects would not have been suitable for the kinds of data to which the theory was to be applied.** The approach taken is, however, considered adequate for two reasons. First, the Community Input clusters were to be used as controlling variables, not as strictly independent or dependent variables; that is, their joint effect in controlling the relationship between School Mediation and Student Output was of interest rather than the exact structure of their effect, and for the purpose of theoretical and empirical analysis they were to be treated only as a set and not separately. Second, it was assumed that by making a sufficiently complex description of general community qualities, most of the significant dimensions which differentiate communities would be included. It can hopefully be said that the more essentially causal dimensions are thereby transformationally related to the ones encompassed by the present clusters. If they are not, then they do not differentiate between communities, and they are not important in the analysis. In empirical terms, it is assumed that the present clusters yield measures which account for the same variation in School Mediation and Student Output as would a more causal set, and therefore the controlling effect is the same.

The "social area analysis" approach adopted for defining clusters led to the adoption of the five main groups of clusters which are explained in the following paragraphs. The studies cited in Chapter 2 of the

*For example, what causes a particular mental trait?

**For example, mental test data were not available for entering students, only for graduating students.

report provided initial inspiration for devising the clusters. However, an attempt was made to associate purely analytic constructs with their potential effect on student growth--i.e.,on input and as continuing influence. Each cluster is defined in terms of what a high value on the underlying scale would mean, and notes are made concerning differentiation that could be made between aspects of the defined construct. Examples of possible indicators for the constructs are presented in Section B of this chapter. The presentation of the clusters is summarized in the upper left box of the foldout.

The first group of clusters defined represent the Population Dynamics of communities. The population across a state is not uniform: age distributions, rates and direction of population change, and densities of population vary. Such dynamics of population determine in part the quality of life in an area: the distribution of age affects the influence of adults on youths; sudden increase in population may cause overcrowding in elementary schools; low density of population may reduce peer group sizes. Three aspects of Population Dynamics are identified and defined as clusters in the theory.

> A1. Growth. A high value corresponds to a large recent increase in population; differentiation may be made according to cause of growth such as high birth rate and immigration.
>
> A2. Youth. A high value corresponds to a concentration of population in the low (e.g., under 25) age group; differentiation may be made according to the modal qualities of the age distribution, such as a high teenage population or a large retirement community.
>
> A3. Density. A high value corresponds to a densely populated area; differentiation may be made according to evenness of area population distribution.

The socioeconomic milieu of an area is described by the Population Characteristics clusters. Three major dimensions are recognized here: the rural/urban character of an area and two aspects of the socioeconomic level of individuals. The occupational aspirations of students, for example, are especially correlated with the rural/urban distinction: farming areas produce farmers. And, for example, cognitive ability correlates with (but is not caused by) socioeconomic level. Rural/urban character is represented by the first cluster. Because socioeconomic level for an area is a distribution of individual levels, three clusters for it are defined.

> B1. Rural/Urban. A high value corresponds to a high amount of urban-type life; differentiation may be made according to the process of urbanization.
>
> B2. Socioeconomic Level. A high value corresponds to a high average socioeconomic status of individuals across the area; differentiation may be made according to the types of socioeconomic class.

B3. Socioeconomic Variation. A high value corresponds to a high amount of variation in socioeconomic status of individuals across the area; differentiation may be made according to modal properties of the socioeconomic distribution--that is, to social stratification.

The next group of clusters concerns the social processes of a community and is called Socialization Characteristics. The attitudes of students and their general responsiveness to education are influenced by the social processes of their community. For example, there may be a sense of resignation associated with the social life found in poverty-stricken areas. There may be a sense of hope in areas where there is much community activity.

C1. Social Activity/Involvement. A high value corresponds to a high amount of social activity and involvement on a community-wide basis; differentiation may be made according to the type of activity, such as community improvement or community recreation.

C2. Social Mobility. A high value corresponds to ease of social mobility; differentiation may be made according to its applicability to different social classes and according to the change in mobility across time.

C3. Poverty Culture. A high value corresponds to a large number of poverty-culture conditions; differentiation may be made according to the kinds of poverty-culture conditions such as family incohesiveness and alienation from community.

By Economic Structure is meant the arrangement of the total economic situation in a community. This helps determine the particular occupational and, hence, educational aspirations of students and is an important practical complex of dimensions along which communities vary. Four clusters are identified:

D1. Wealth. A high value corresponds to general economic affluence in a community; differentiation may be made according to peculiarities in the distribution of wealth, to circulation or accumulation of wealth, and to wealth in public or private sectors of the economy.

D2. Industry. A high value corresponds to the existence and vitality of industrial operations in a community; differentiation may be made according to whether industry is light or heavy and according to the number of different industries.

D3. Trade. A high value indicates a large amount of trade activity; differentiation may be made according to whether trade is wholesale or retail and according to the kind of sales force present.

D4. Farming. A high value indicates a large involvement in farming; differentiation may be made according to type of farming products (dairy, cattle, etc.) and to how modern is the operation.

The final group of clusters concerns the Educational Resources of a community. The dimensions indicate the continuing influence of the general educational atmosphere of a community. The interest of students in education and their prior goals and achievements in education are determined in part by the educational standards, achievements, and interest of the community members.

E1. Archive Usage. A high value corresponds to high usage of various formal and informal archives of knowledge; differentiation may be made according to the types of archives available--libraries, museums, etc.

E2. Educational Involvement. A high value corresponds to involvement and interest in education of various forms of a large part of the population; differentiations may be made according to the kind of education--academic, vocational, avocational--and according to the direction of involvement--participation, administration, encouragement.

E3. Educational Attainment. A high value corresponds to high average level of education or skill; differentiations may be made according to kind of education and to stratification of attainment.

School Processes as Mediation

A student enters high school with certain abilities and aspirations. He has a personality produced in home, community, and elementary school which can be described in terms of his abilities, aspirations, and emotional outlook.* When a student leaves high school he may be again viewed in terms of his personality. The change in the personality, the growth of the student, is partly an effect of the community (as defined before). This mediation by the school is to be analyzed in the theory in terms of the primary causes of the effect of the school in student growth. The processes which go on in the school and which directly cause effects in the student are to be defined. For this purpose the total school environment must be considered, including social and cultural processes as well as the more formal teaching processes. The processes to be defined are essentially individualistic; they involve the interaction of a single student with his environment. Since the measurement is to be made in terms of schools, separate mean-level and variation clusters

*See the characterization of Student Output for a more detailed categorization.

are defined for each process considered. This approach was explained at the beginning of the section.

The School Mediation complex is divided into two groups of factors which are intended to be all-inclusive in terms of what happens to students in school. A student has certain stimuli directed at him and he obtains certain satisfactions from the school environment; i.e., the mediation processes are generally considered to be those of stimulation or those of satisfaction.

<u>Stimulation</u>. In the school environment a student is bombarded with various stimuli such as facts, ideas, opinions; the stimuli converge continuously and often simultaneously upon him. In the theory the stimuli are classified according to a three-way design and amount of stimulation is the quantity to be scaled. Some sort of classification is necessary in order to make possible the evaluation of the different modes of stimulation achieved in different kinds of schools. The first classification is the mean level versus variation distinction, already mentioned, which adjusts individualistic to wholistic measurement. The second classification concerns the substances of stimuli, and the third concerns the structure of stimuli. A stimulus is considered to be a conveyance of information (substance) directed toward the student in a certain manner (structure). For example, in a language laboratory, the vocabulary and grammar of a language are the substance of the stimuli directed at the student; and recorded criterion voices, recorded student response, and immediate teacher feedback make up the structure of the stimuli. The average time in the laboratory across students is the mean-level stimulation, and the variance of time across students is the variation in stimulation. The stimulation clusters then follow a design. In the

Mean-Level
Variation x Substance x Structure

following paragraphs, the levels within substance and within structure are explained and thus the clusters are defined. They are diagrammed at the top of the middle box in the foldout at the end of this section.

The substance of stimulation is divided into nine levels. The first five correspond to the standard formal academic areas:

<u>F1. English</u>.

<u>F2. History</u>.

<u>F3. Science</u>.

<u>F4. Mathematics</u>.

<u>F5. Foreign Language</u>.

The correspondence to the formal administration classification is, it should be emphasized, imperfect; the concern of the theoretical classification is for the information being conveyed. While within administrative classifications, teaching English composition may be one purpose of history courses, in the present classification the concern is purely for the substance of the stimulation, so such crossover is explicitly disregarded--composition taught in a history course is still English stimulation. Three other levels also have corresponding formal course work.

F6. Culture. (Music, art, drama.)

F7. Vocational Training. (Both commercial and trade.)

F8. Physical Education. (Including interscholastic competition.)

The last level i· composed of substance of stimuli not covered in the other eight levels.

F9. Auxiliary. (Including exceptional education, guidance and counselling, remedial reading, etc.)

The substance of a stimulus is directed at the student in a certain way; in other words, the mediation occurs in a certain fashion; a stimulus has structure. The essential differentiation among structures of stimuli is viewed in terms of social interaction. Three levels are defined.

G1. Primary Social.

G2. Secondary Social.

G3. Asocial.

When a student discusses current events with his friends, he receives stimulation with a Primary Social structure. When he listens to a large group lecture from a teacher, he receives stimulation with a Secondary Social structure. When he sees a film, the stimulation has Asocial structure. Stimulation of rimary Social structure is usually from peers; Secondary Social structure from teachers; Asocial structure, equipment.

Now a particular cluster for stimulation may be specified as a combination of either mean-level or variance, plus one of the nine levels of substance, plus one of the three levels of structure. For example, the average time spent in a language laboratory might be an indicator for the cluster:

Mean-Level / Foreign Language / Asocial.

A clique of students discussing a novel might be an indicator for

Variation / English / Primary Social.

The particular classification of structure of stimulation was designed so that an apparently important effect of school size might be detected. In a large school (or in any large organization) the quality of social interaction is different from that in a smaller school. In a small school everyone knows everyone; in a large school there may be strangers in a single classroom. In a small school a teacher may see a particular student for several periods in a day, while in a large school this may be unlikely. Consequently, the social atmosphere in which teaching and learning take place is different, and this may be expected to account for much of the difference between large and small schools. Also large and small schools are known to have different possibilities in course offerings and in sectioning classes, so that the substance of the stimulation may be expected to vary.

<u>Satisfaction</u>. A student derives or fails to derive certain pleasures in school; i.e., the school environment is the setting in which some experiences (for example, receiving stimuli) are intrinsically pleasurable or not pleasurable. While stimulation refers to activity in which a student engages, satisfaction describes the state-of-being a student enjoys at a particular point in time. Aspects of satisfaction are categorized according to the kind of individual pleasure.

<u>H1. Intellectual.</u> Acquisition of knowledge is a goal of society; learning is considered intrinsically desirable. Learning a new concept or fact, discovering a particularly interesting piece of literature, or successfully completing a scientific investigation might be occasions for intellectual satisfaction.

<u>H2. Sensory</u>. Physical activity may be intrinsically pleasurable. Participation in sports or dance courses might yield sensory pleasure. Note that sensory pleasure is independent of any information being obtained about physical activity; it involves playing a game, but not learning the rules.

<u>H3. Social</u>. There may be intrinsic pleasure in interaction with others. Friendships with peers or with teachers give pleasure: primary social interaction may be intrinsically pleasurable. Formal recognition, such as being elected a class officer, may yield social satisfaction from secondary interaction.

These three aspects of satisfaction apply to individual students. Consequently, it is necessary, as with stimulation, to form clusters according to a design. The design for satisfaction is

Mean-Level Variation	x	Satisfaction

Since there are three levels defined for satisfaction, six clusters are defined in all. For example, the presence of both students who participate in no sports and students who participate in all sports might be an indicator for

Variation / Sensory Satisfaction.

This is a two-way design and is diagrammed at the bottom of the middle box in the foldout.

The Student as Output

As explained above, the qualities of students upon entering high school are theoretically and empirically described only indirectly through the general characteristics of the community. The qualities, though not explicitly defined before, were assumed to be mediated in the high school environment. The kinds of qualities of students being considered are the same for input as for output: the clusters would ideally be parallel, but the values of the scales are changed by high school mediation and continuing community influence. Student Output is, of course, a complex construct. No single indicator--such as percentage of graduates going on to college or percentage getting good jobs--adequately describes the complexity of results, for there must be concern for the total life style of the student after high school. The major groups of output clusters defined in the theory are Abilities, Aspirations, and Emotional Outlooks. The three groups of clusters are related more or less to the established goals of secondary schools. This relation was considered desirable because the theory is to apply to the practical problem of what administrators affect when they manipulate such factors as school size.

Again, all defined outputs must be considered in mean-level and variation effect clusters. The outputs are characteristics of students, and the distributions of the characteristics are to be described. The clusters are listed in the lower right box in the foldout.

Abilities. Each student has certain ability characteristics. An ability may be defined as intrinsic or extrinsic capacity or attainment. Note that achievement, in the sense of knowledge or skill, and such extrinsic characteristics as socioeconomic status are regarded as abilities. Three clusters of abilities are identified.

I1. Cognitive. (Including intrinsic knowledge, skills, and mental traits.)

I2. Psychomotor. (Including physical capacity and attainment.)

I3. Social. (Including capacity for interaction and social standing.)

The formal academic goals of a school have representation in the Abilities clusters.

Aspirations. Students have aspirations, which are conscious or unconscious desires for long- or short-term activities. Passing next week's examination is an aspiration, and so is becoming a policeman. An aspiration may be unrecognized, such as a desire to get a higher grade than a friend. In the theory there are three clusters related to aspirations.

J1. Vocational. A high value corresponds to aspiration to a high status or an appropriate job.

J2. Avocation. A high value corresponds to the development of an interesting hobby.

J3. Social. A high value corresponds to the aspiration to social interaction, friendship, and leadership.

Community and especially family influences are very strong in developing aspirations. Also, the aspirations of a student may or may not be consonant with the opportunities of the community or the economics of the family. A student may aspire to a job which is not available to him in his community. Or a student may aspire to college when his family cannot afford it. Thus when determining the components of aspirations, it is important to consider their relation to the abilities of the community, family, and even the student himself.

Emotional Outlooks. The attitudes of a student are crucial in determining his success in life and, earlier, in school. Of particular interest with respect to school goals and facilities is the general receptiveness of a student to growth. A student may be hostile to learning, or he may be distracted from study by emotional problems. Three clusters of emotional outlooks are identified:

K1. Autonomy. A high value corresponds to freedom and ability to make choices and is an essential ingredient of emotional adjustment.

K2. Distraction / Inhibition. A high value corresponds to complications in the responsive systems of the mind as, for example, through having several simultaneous stimuli, or through having blocking reactions.

K3. Social. A high value corresponds to the emotional ability to undergo social interaction and to general adjustment and happiness.

The Emotional Outlooks of the students are considered important inputs for their further education.

School Manipulation

All the clusters describing the qualities, characteristics, and processes involved in the causal chain have now been presented. School manipulation variables are, as mentioned, outside this causal chain and are considered to affect the Student Output clusters only indirectly in their effect on School Manipulations. The focus of the present study is on school size (enrollment) as the single school manipulation variable. To complete the presentation of the theory, other manipulation variables to which the theory could be applied are listed in the upper right box of the foldout.

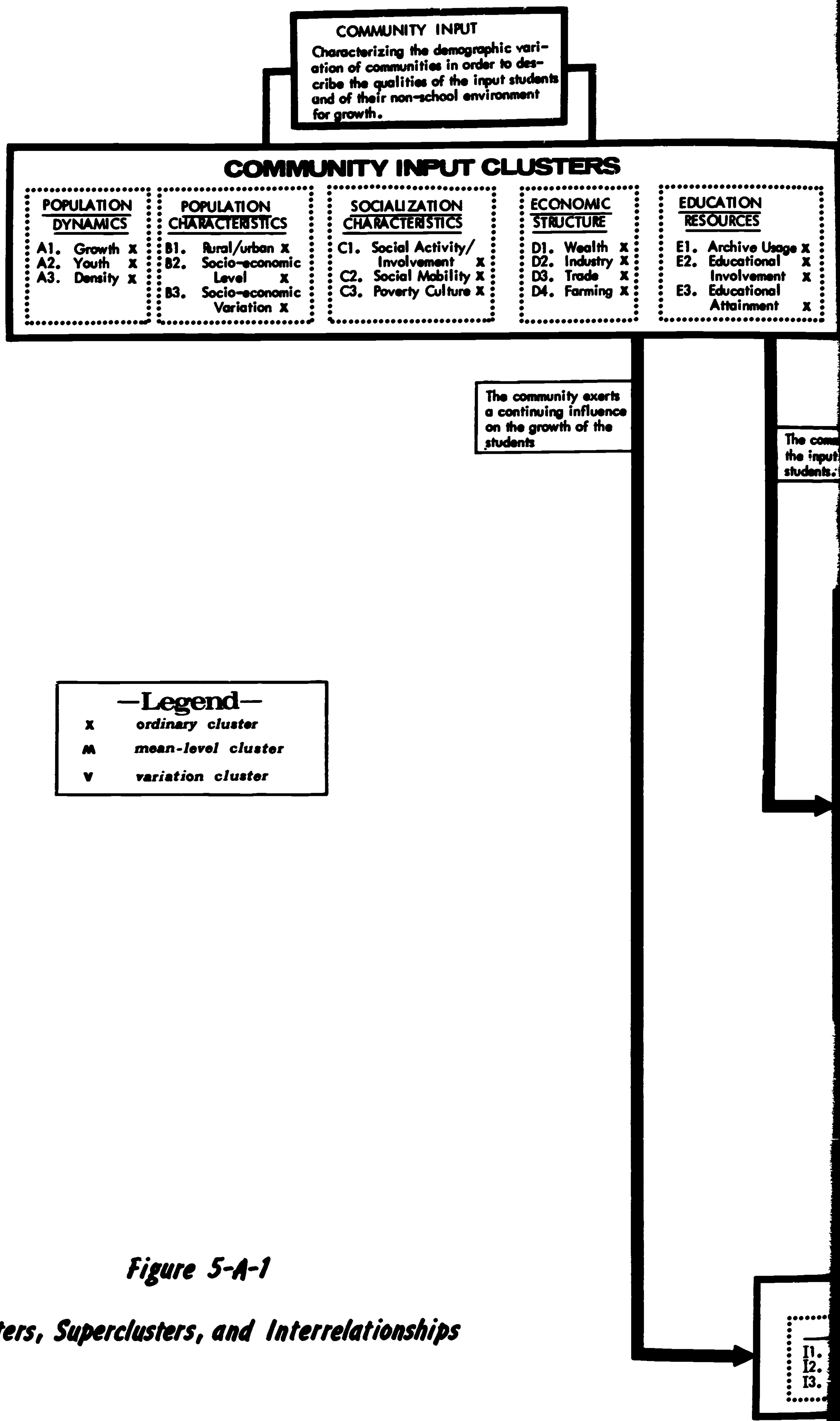


Figure 5-A-1

Clusters, Superclusters, and Interrelationships

SCHOOL MANIPULATION
Encompassing the design factors such as size which can be manipulated directly, but which do not directly affect student output.

EXAMPLE OF SCHOOL MANIPULATION VARIABLES

1. School Size (enrollment)
2. School Size (plant)
3. Administrative Complexity
4. Design Factor (e.g., kinds of chairs)
5. Student/teache- Ratio
6. Textbooks Used
7. Hot Lunch Program

ommunity determines
put qualities of the
nts.

The School Manipulation variables affect the school mediation processes.

SCHOOL MEDIATION
Consisting of processes and forces directed at students in high school and of their concomitant reactions.

SCHOOL MEDIATION CLUSTERS

STIMULATION

SUBSTANCE	STRUCTURE: Primary Social G1.	Secondary Social G2.	Asocial G3.
	MV	MV	MV
F1. English	MV	MV	MV
F2. History	MV	MV	MV
F3. Science	MV	MV	MV
F4. Mathematics	MV	MV	MV
F5. Foreign Language	MV	MV	MV
F6. Culture	MV	MV	MV
F7. Vocational Training	MV	MV	MV
F8. Physical Education	MV	MV	MV
F9. Auxiliary			

SATISFACTION

H1. Intellectual MV
H2. Sensory MV
H3. Social MV

The school mediation helps transform student input to output.

STUDENT OUTPUT
Consisting of the qualities of the students' output from high school in terms of their change from input status and their receptiveness for future growth.

STUDENT OUTPUT CLUSTERS

ABILITY

1. Cognitive M V
2. Psychomotor M V
3. Social M V

ASPIRATION

J1. Vocational M V
J2. Avocational M V
J3. Social M V

EMOTIONAL OUTLOOKS

K1. Autonomy M V
K2. Distraction M V
K3. Social M V

B. Ideal Measurement and the Availability of Data

The relationship between high school size and educational productivity was conceptualized in the previous section. This conceptualization was needed in order to define the data analytic operations. It was also needed to evaluate the adequacy of the data bank contents for investigating the substantive relationship between school size and non-dollar costs. The theory is detailed because the relationship is complex. It was essential to be able to describe the various effects of size and to determine the mechanisms through which the effects are transmitted. In particular, the social characteristics of the community and school environments and the social consequences of size are emphasized. It is postulated that the major difference between large and small schools lies in their social qualities.

The Iowa Educational Information Center and the U.S. Census Bureau data banks were established for purposes other than the present research. The Iowa data bank grew out of the school and student accounting service established on a statewide basis. The U.S. Census Bureau data bank was established originally for electoral redistricting and has grown as new demands for information were made. From the viewpoint of this research, neither source makes adequate measurement of the social characteristics and processes in its domain.

This section contains information on how ideally measurement would be made for the clusters of the theory. The information is divided into sub-sections, one for each of the superclusters--Community Input, School Mediation, and Student Output. While the paucity of data has led to a partial redefinition of the clusters, an argument for the superiority of the original clusters is made.

Measuring Community Input

An example of an ideal indicator for each of the Community Input clusters is given in Table 5-B-1. The Census Bureau data were found to be adequate; several ideal measurements were achieved and many reasonable alternate indications were found. However, three shortcomings in Census Bureau data must be noted.

First, with respect to educational data (1965-1966), the census data (1960) are older. No research study can use new census data except in the decennial year. A special census would have been enormously expensive and using older school data would have been impossible since educational data banks were not in existence in 1960. Some 1960 figures, such as population, are fairly accurate substitutes for 1965 figures; i.e., the correlation across school areas between 1960 and 1965 populations is close to one. In the areas where there are discrepancies, it must be hypothesized that social and economic changes have taken place. For example, if a community's population has increased greatly from 1960 to 1965 relative to other communities, it is likely that some immigration has occurred and that urbanization is taking place. Such changes are

TABLE 5-3-1

IDEAL INDICATORS FOR COMMUNITY INPUT

Cluster	Indicator
A1. Growth	% population increase 1960-1965
A2. Youth	% of population less than 20 years old
A3. Density	Population per square mile
B1. Rural/Urban	% of population living on farms
B2. Socioeconomic Level	Mean family income
B3. Socioeconomic Variation	Variance of family income
C1. Social Activity/Involvement	% of adults belonging to a social club
C2. Social Mobility	Difference between parent and child occupation
C3. Poverty Culture	Unemployment rate
D1. Wealth	Per-capita personal wealth
D2. Industry	Per-capita industrial business worth
D3. Trade	Per-capita retail sales
D4. Farming	Per-capita farm worth
E1. Archive Usage	Per-capita book circulation
E2. Educational Involvement	Per-student PTA attendance
E3. Educational Attainment	% of adults who are college graduates

very much related to the social and economic environment of a community. In particular, they must be considered related to school size and non-dollar costs. A student who has recently moved into a community makes friends. The processes involved may be different in large and small schools. This reasoning can easily be extended to other variables. It implies that measurement error in using old census data is likely to correlate with the essential variables of the analysis and, therefore, is likely to cause distortion in the results.*

The second major problem with census data concerns the measurement of the social characteristics of a community. For the purpose of the present research, measurement would ideally be made of social processes as they occur in a community. But the Census Bureau does not make social inquiries, such as asking persons to what organizations they belong. In defining the social climate of a community--the quality of the social environment and life of its occupants--the absence of direct social inquiries is a limitation. With respect to the present problem of examining the relationship between school size and non-dollar costs, defining this climate is important: the social life of the community at large overlaps with the social life of the school. In the present theory, the social life of the school is considered the crucial intervening variable between size and non-dollar costs.

The third point which should be made about census data is that the census is an inventory of individual characteristics. It is not intended to describe the institutions of a community. This makes it difficult to measure the economic structure of a community. Using census data, economic structure must be inferred from the occupations of the people of the community, and this probably does not yield an adequate description of the dynamics of the economic structure. Similarly, public institutions--such as churches, libraries, and schools--must be described indirectly through certain characteristics of individuals. Essentially no information is available on the avocational interests satisfied by institutions such as social and service clubs. This is a limitation in terms of the need according to the theory to define the continuing effect of the community on the social and educational growth of students in high school.

Measuring School Mediation

There are difficulties in measuring from educational data bank information defined for School Mediation by the clusters. To overcome these difficulties, a simplification of the cluster scheme was adopted for performing the data analysis; this simplification is presented in the next section. Some further notes are presented first on why the complexity of the original scheme should be considered necessary for accurate description of the process of mediation by the school.

*Using 1950 census figures for the purpose of measuring and projecting such change would have been prohibitively expensive. It would probably also have yielded inadequate projections since many school areas are small geographic areas capable of rapid change.

In the original scheme, there is a three-track design of clusters concerning the Stimulation that students receive in school. The first track in the design is Substance. Its levels correspond to nine major subject matters in the curriculum of a high school. Larger schools usually offer more courses; e.g., they offer more advanced courses in subjects such as science and foreign language. Therefore, it is important in relating school size to student output to be able to consider the differential effect of program offerings--i.e., of Stimulation according to the various subject matters. The second track in the design is Structure, and its levels correspond to the mode of stimulation--by primary social interaction, secondary social interaction, and by asocial means. The present theory places emphasis on the social differences between large and small schools. The social atmosphere in a classroom differs qualitatively according to school size. For example, the students and the teacher may not know one another as well in a large school; in a small school there may be more school-wide association. In a homogeneously grouped classroom in a large school, there will be less variation in ability and attitude, causing less conflict than in a heterogeneous classroom of a small school. Therefore, levels of Structure are considered necessary for description of the essential differences between large and small schools in terms of how learning takes place. The third track is distinguished by mean level and variation of stimulation. This distinction is necessary since the amount of stimulation for a school is really a distribution of the amounts received by the individual students. In summary, the original clusters for describing Stimulation were necessary, and the lack of data adequately to measure them is a stated limitation.

Satisfaction has clusters defined according to a two-track design in the original scheme. The levels of the first track are intellectual, sensory, and social satisfaction. These levels represent an attempt to define the emotional qualities which may be degraded in large and densely populated social settings. In particular, it is suspected that large schools may cause students to feel alienated and to forget academic goals because of emotional problems. The distinction between the three kinds of satisfaction seems necessary in order to determine the dimensions of emotional attitude in large and small schools. The second track corresponds to mean level and variation in amount of Satisfaction. Again, the amount to be measured is the distribution of amounts for individual students.

In terms of designing further research into the question of school size and non-dollar costs and in terms of further defining the insufficiency of the available data, it is useful to consider what kinds of indicators ideally should be available for satisfying the original scheme. A list of indicators is presented below:

FI-GI-M. Mean Level of English Primary Social Stimulation

1. mean number of hours students spend discussing literature in small groups

2. homogeneity of student ability in English classes

FI-G2-M. Mean Level of English Secondary Social Stimulation

1. mean number of hours a student attends lectures about literature

2. mean amount of friendship of students toward English teacher

FI-G3-M. Mean Level of English Asocial Stimulation

1. mean number of hours a student spends reading assigned books

2. mean amount of blackboard usage in English classes

HI-M. Maan Level Intellectual Satisfaction

1. mean number of unassigned books a student reads

2. proportion of students receiving honor awards

H2-M. Mean Level Sensory Satisfaction

1. mean lighting power of the classrooms
2. proportion of students engaged in extracurricular athletics.

H3-M. Mean Level Social Satisfaction

1. mean number of hours students spend in private conversations in school

2. number of school social events.

Examples are given for mean-level primary, secondary, and asocial structure stimulation in English and for mean-level intellectual, sensory, and social satisfaction. Such data clearly require special data collection, since it would serve no auxiliary practical function, such as school accounting, in a data bank. Yet, it would provide much more exact explanations of the differential effects of school size on student output.

There is another problem in dealing with the Stimulation clusters as originally defined. The subject areas for Substance were explicitly specified as referring to kinds of substance, not to administrative units such as classes. English Literature as taught in history classes should be tabulated as English stimulation. The amount of crossover in subject matter may be expected to vary according to size of school; in a large school, teachers may be more specialized in what they teach. This distinction can be lost in the data available in education data banks and the result is distention.

Measuring Student Output

The adequacy of the educational data bank for measuring the clusters defined for Student Output is mixed. Precise measurement of Cognitive Ability is possible; in fact, the Cognitive Ability clusters may be subdivided. However, there are no measures of Psychomotor and Social Ability and there are not sufficient data to maintain the distinctive clusters concerning aspects of Aspiration and Emotional Outlook. Because of the abundance of cognitive ability data and the paucity of other data, a revised scheme of Student Output clusters was designed. It is presented in the next section. Here, the implications of the abundance and paucity are discussed.

Examples of ideal indicators for the clusters as originally designed are presented below. Only the mean-level clusters are given.

I1-M. Mean-Level Cognitive Ability

1. mean composite score on ability-achievement test
2. proportion of students planning to attend college

I2-M. Mean-Level Psychomotor Ability

1. mean score of students on a test of physical fitness
2. proportion of students regularly participating in sports

I3-M. Mean-Level Social Ability

1. mean number of friends of a student
2. proportion of students attending social events

J1-M. Mean-Level Vocational Aspiration

1. mean prestige score of vocational aspiration
2. mean discrepancy between student's occupational aspiration and parent's occupation

J2-M. Mean-Level Avocational Aspiration

1. mean number of hours spent per week on a hobby
2. mean number of hours spent per week watching television

J3-M. Mean-Level Social Aspiration

1. proportion of students aspiring to a social occupation
2. proportion of students planning to get married soon

K1-M. Mean-Level Autonomy

1. proportion of students who are certain of their plans
2. proportion of students with paying jobs.

K2-M. Mean-Level Distraction

1. correlation between mark-point average and composite test score
2. average number of days absent

K3-M. Mean-Level Social Outlook

1. proportion of students with girl (boy) friends
2. proportion of students who report that they are generally happy.

The variety of cognitive ability data is due to the coding in the data bank of all subscales of the Iowa Test of Educational Development for all seniors. In redefining the Cognitive Ability clusters, various groups of subscales are clustered. This operation was motivated by recent work by Bock and Wiley (personal communication) which suggests, on the basis of a study in the Chicago area, that schools produce characteristic patterns of subscale responses on ability and/or achievement tests. In terms of the present research, it is very relevant to ask whether large or small schools produce students with distinctive patterns on test subscales--i.e., is there a differential effect of school size on verbal and quantitative scales?

On the other hand, the paucity of data on the other facets of output impose a limit on the detail in which the effect of school size can be specified. The emotional makeup of a student, his plans, his interests, his health, and his social qualities are admitted targets of secondary education. In its most obvious form, this makeup is manifested in the teaching and grading of "citizenship". A large part of a school's efforts are aimed at these non-academic concerns. Yet, while for academic areas, there is an elaborate attempt to evaluate both the system and the individual, there is essentially no recorded information for evaluating the non-academic areas. To obtain such data would, of course, be technically difficult--and perhaps not entirely desirable in terms of long-range standards and goals of society--but its absence certainly makes difficult the evaluation of school systems and of school size.

C. Clustering of the Available Data

In Section A of this chapter, a theoretical framework was presented for studying the relationships between high school size and educational productivity. In Section B, notes were presented on what kinds of data ideally would be used to perform analysis within that framework and within the limitations of the data available from the Iowa and U.S. Census Bureau Data Banks. It was not expected that the available data would satisfy a complex a priori theoretical scheme. Yet, it was considered essential to develop the scheme and to work as far as possible within it. This process was described in Chapter 2.

This section contains descriptions of clusters of indicators actually determined and computed from the data samples. As explained in the previous section, certain general revisions of the original clusters had to be made. In addition to these general revisions--in School Mediation and Student Output--numerous minor revisions were made. The "input" of a cluster is defined according to its indicated composition. When indirect or approximate indicators were available for a cluster, the meaning of the cluster is changed. It should be emphasized that these changes do not represent error in the sense of random measurement error, but rather error in the sense of change in the face value or validity of the clusters.

The description is given in three parts, one for each of the superclusters--Community Input, School Mediation, and Student Output. For each supercluster, the revisions necessitated by the availability or unavailability of data are presented. The indicators are then computed and the revisions in their assignment given. One point should be made on the definition of variables. There are three types of school organizations represented in Iowa: six; four; and three-year high schools. To make the data compatible, usually only data concerning students in grades 10-12 inclusive were used.

Community Input

As noted in the previous section of this chapter, there were sufficient data in the Census Bureau data bank to compute indicators for all of the pre-defined community input clusters. It was not necessary, therefore, to make changes in the theoretical scheme. It should be reiterated that there are several shortcomings in the use of census data. First, because of the history and nature of the census, the data necessary for constructing ideal indicators were not available; i.e., the kinds of community social processes which are considered important for the measurement of the effect of school size on educational productivity could not be directly indicated, but had to be inferred at a certain loss of accuracy. Second, the census data are old with respect to the school data used. The changes which may have occurred in communities certainly would be crucial to the question of school size. The fact that a community has changed recently would have ramifications for the social processes operating in the community and for the consequent social characteristics of the students and the success of their performance in large or small schools.

Population Dynamics. The Population Dynamics clusters consist of the following indicators:

A1. Growth

A1.1 Birth rate: persons less than one year old per 1000 population

A1.2 Age of residence: percent of residential structure built after 1950

A1.3 Immigration: percent of population living outside the county in 1955

A2. Youth

A2.1 Infants: percent of population less than 6 years old

A2.2 Children: percent of population 6-19 years old

A2.3 Young males: percent of population males 20-29 years old

A2.4 Young females: percent of population females 20-29 years old

A3. Density

A3.1 Structure density: housing units per residential structure

A3.2 Unit density: persons per housing unit

A3.3 Room density: persons per room in housing units.

A high amount of growth is indicated by 1) a high birth rate, 2) new homes, and 3) a high rate of immigration. These are indirect measurements, since comparative figures from earlier years could not be obtained. Youth has four indicators corresponding to four segments of that population: 1) infants, 2) children and teenagers, 3) young adult men, and 4) young adult women. The differentiation between young adult men and women was made because of the possibility of different rates of post-school migration for young men and women. The three indicators provided for density are measures of housing density in three respects: 1) multiple-unit structures, 2) crowded units in terms of the number of people residing in them, and 3) crowded units in terms of the number of rooms. Since the geographic areas of the communities were not obtained, no indicator directly measuring land crowding was constructed.

Population Characteristics. The Population Characteristics clusters consist of the following indicators:

B1. Rural/Urban

B1.1 Urban Residence: percent of population with urban residence

B1.2 Rural Non-Farm Residence: percent of population with rural non-farm residence

B1.3 Agriculture Employment: percent of labor force employed in agriculture

B2. Socioeconomic Level

B2.1 Income Level: mean of family income, measured on a 13-point scale from 1 = $0-999 to 13 = $25,000+

B2.2 Education Level: mean of education for persons 25 years and older, measured on an 8-point scale from 1 = no schooling to 8 = 4 or more years of college

B2.3 Occupation Level: mean of occupation for employed males, measured on an 11-point scale from 1 = laborer to 11 = professional

B3. Socioeconomic Variation

B3.1 Income Variation: variance of family income, measured on a 13-point scale from 1 = $0-999 to 13 = $25,000+

B3.2 Education Variation: variance of education for persons 25 years and older, measured on an 8-point scale from 1 = no schooling to 8 = 4 or more years of college

B3.3 Occupation Variation: variance of occupation for employed males, measured on an 11-point scale from 1 = laborer to 11 = professional.

The rural/urban character of a community is indicated by 1) the proportion of the population having urban residence, 2) the proportion of the population having rural non-farm residence, and 3) the proportion of the labor force employed in agriculture. The Census Bureau defines three types of residence: urban, rural non-farm, and farm. The rural non-farm corresponds to rural homes not on farms and to residences in hamlets. The socioeconomic situation in a community is represented by the standard trilogy of 1) income, 2) education, and 3) occupation. Because the total community situation is construed to be a distribution of the socioeconomic levels of the members of the community, there is a pair of clusters. One describes the mean-level status and one describes the variation in status. The income data are for families. The education data are for adults 25 years and older who usually have completed their formal education. The occupation data are arranged according to an approximate prestige rating of the census, ranging from laborer to professional.

Socialization Characteristics. The Socialization Characteristics clusters consist of the following indicators:

C1. Social Activity/Involvement

C1.1 Social workers: social, welfare, and recreation workers per 1000 population

C1.2 Public employees: percent of labor force employed by the public

C1.3 Female employees: percent of adult females who are employed

C2. Social Mobility

C2.1 Population heterogeneity: percent of population non-white, foreign born, or of foreign stock

C2.2 Population mobility: percent of population living in a different house in the same county in 1955

C3. Poverty Culture

C3.1 Low income: percent of families with income less than $2,000

C3.2 Poor housing: percent of residential housing units which are deteriorating, dilapidated, or which lack any facility

C3.3 Unemployment: percent of able males who are unemployed.

As stated above, the Census Bureau does not collect data which are descriptive of the social processes of a community. Therefore, the indicators for social activity/involvement are derivatives. Three indicators were computed: the first is an index of the number of persons in the community whose occupations indicate direct social activity; the second, proportion of the labor force employed by the public may be related to the extent of community responsibilities; the third, proportion of adult females who have jobs, may be related to the social activity of women in the community. The assumption is that women who work outside the home have more opportunity for social contact. The indicators for the social mobility cluster are also derivatives. A high heterogeneity of the population may indicate the presence of social change through assimilation or conflict; the housing mobility of the population may indicate changing socioeconomic situations. The indicators chosen for the poverty-culture cluster are versions of three standard evidences of poverty: 1) low income, 2) poor housing, and 3) unemployment.

Economic Structure. The Economic Structure clusters consist of the following indicators:

D1. Wealth

D1.1 Property value: mean residential property value, measured on a 10-point scale from 1 = $0-4,999 to 10 = $35,000+

D1.2 Rent level: mean gross rent, measured on a 13-point scale from 1 = $0-19 to 13 = $200+

D1.3 Automobile availability: number of automobiles available per housing unit

D2. Industry

D2.1 Industry employment: percent of labor force employed in industry

D2.2 Managerial employment: percent of labor force employed as managers

D2.3 Clerical employment: percent of labor force employed as clerks

D3. Trade

D3.1 Retail employment: percent of labor force employed in retail establishments

D3.2 Service employment: percent of labor force employed in personal and business service

D3.3 Sales employment: percent of labor force employed as sales workers

D4. Farming

D4.1 Family farm help: unpaid family farm workers per farmer

D4.2 Paid farm help: paid farm laborers per farmer

D4.3 Food industry: percent of labor force employed in the food industry.

The indicators for the wealth clusters were aimed at measuring personal wealth in terms of: 1) residential property value, 2) rent level, and 3) availability of automobiles. The wealth measures were related to personal wealth rather than corporate wealth, since census data are collected for individuals. The industry cluster is indicated by 1) employment in industry, 2) employment in managerial positions, and 3) employment in clerical positions. The presence of a business bureaucracy is included in this cluster by distinguishing managerial and clerical positions. The trade cluster is indicated by 1) retail, 2) service, and 3) sales employment levels. Business and personal services are combined in this cluster, although they are separate census classifications. The indicators for the farming cluster are 1) number of family farm workers, 2) number of farm laborers, and 3) percent of labor force in the food industries. They are intended to index the farming economy.

Educational Resources. The Educational Resources clusters consist of the following indicators:

E1. Archive Usage

E1.1 Librarians: librarians per 1000 population

E2. Educational Involvement

E2.1 School enrollment: percent of population enrolled in school

E2.2 School employment: percent of labor force employed in educational services

E3. Educational Attainment

E3.1 College attainment: percent of persons 25 years and older with some college education

E3.2 Professional employment: percent of labor force employed in professional occupations.

Only one indicator could be found in the census data to index the amount and nature of usage of archives in communities, and that was the number of librarians per 1000 population. The educational involvement cluster has as indicators 1) the percent of the population enrolled in school, which reveals the degree to which families in the community are directly connected with education, and 2) the proportion of the labor force employed in educational service. No further indicators could be found which would show the amount of informal involvement of the community in its education processes. The educational attainment cluster has two indicators: 1) the amount of college education of adults in the community, and 2) the amount of professional employment in the community. As noted earlier, the presence of such persons may be expected to stimulate interest in academic pursuit.

School Mediation

The Iowa educational data bank does not provide adequate indicators for the theoretically defined clusters of the School Mediation supercluster. A revised and simplified theory was therefore developed for the purpose of data analysis. The revised system is diagrammed in Figure 5-C-1; it was derived more from an ad hoc consideration of what data were available than from a re-examination of the theory. The three major areas--stimulation substance, stimulation structure, and satisfaction--are preserved in the revised scheme, but the amount of detail within them and the amount of interaction among them is simplified. The substitution of the revised scheme for the original involves a reduction in descriptive complexity and of the precision with which the variation in student achievement can be assigned to specific processes and qualities in the schools.

In the revised scheme, substance and structure have been separated in the stimulation clusters. In the original scheme, the two aspects were crossed: the substance of the stimulation was differentiated according to the structure of the stimulation. In the revised theory, there are levels of stimulation and levels of structure, but there is not the interaction. Moreover, the number of levels under substance has been reduced to five--only the basic academic subjects are considered--and the differentiation between mean-level and variation effects of substance has been omitted. The single cluster for each of the five levels under substance represents the general amount, quality, and reception of stimulation. The stimulation structure clusters have been reduced to two pairs: mean-level and variation in teacher experience and class heterogeneity. The teacher experience cluster is considered to approximate the amount of interaction taking place between teacher and student, and the class heterogeneity cluster is considered approximate to the amount of interaction taking place among students in the class.

The satisfaction construct is reduced in the revised scheme to two cluster pairs: mean-level and variation in attitude toward study and social interaction. By attitude toward study is meant a student's willingness and desire to learn. By social interaction is meant the quality of interaction a student has with teachers and with peers.

Stimulation

Substance			Structure		
F1.	English	x	G1.	Teacher Experience	MV
F2.	History	x	G2.	Class Heterogeneity	MV
F3.	Science	x			
F4.	Mathematics	x			
F5.	Foreign Language	x			

Satisfaction

H1.	Attitude Toward Studying	MV
H2.	Social Interaction	MV

Figure 5-C-1

Revised School Mediation Clusters

Both these clusters are considered to relate to the amount of satisfaction (as defined in the original scheme) a student obtains from school. Because they are measures of individual qualities, both mean-level and variation clusters are defined.

Stimulation Substance. The stimulation substance clusters consist of the following indicators:

F1. Mathematics

F1.1 Advanced mathematics: percent of students taking an advanced mathematics course

F1.2 Mathematics consistency: correlation coefficient between mathematics marks and MPA's

F1.3 Mathematics/like: percent of students who like mathematics best

F1.4 Mathematics/dislike: percent of students who like mathematics least

F2. Science

F2.1 Advanced science: percent of students taking an advanced science course

F2.2 Science consistency: correlation coefficient between science marks and MPA's

2.3 Science/like: percent of students who like science best

F2.4 Science/dislike: percent of students who like science least

F3. English

F3.1 Advanced English: percent of students taking an advanced English course

F3.2 English consistency: correlation coefficient between English marks and MPA's

F3.3 English/like: percent of students who like English best

F3.4 English/dislike: percent of students who like English least

F4. History

F4.1 Advanced history: percent of students taking an advanced history course

F4.2 History consistency: correlation coefficient between history marks and MPA's

F4.3 History/like: percent of students who like history best

F4.4 History/dislike: percent of students who like history least.

F5. Foreign Language

F5.1 Foreign language: percent of students who are taking a foreign language course

F5.2 Foreign language consistency: correlation coefficient between foreign language marks and MPA's

F5.3 Foreign language/like: percent of students who like foreign language best

F5.4 Foreign language/dislike: percent of students who like foreign language least.

All indicators are based on students in grades 10-12 inclusive only. For each subject matter area, advanced courses are defined as those which are at a higher than required level, such as second year algebra, or which have nonstandard content, such as astronomy. The correlation coefficients are based on all course grades for each subject matter area. The like and dislike percents are based on responses to two questions concerning which subject the student liked best and which he liked least: there are five alternatives other than these five subjects.

Similar indicators are defined for each subject--mathematics, science, English, history (social studies), and foreign language. The first indicator for each subject is the percent of students in grades 10-12 who are taking an advanced course in the subject. This indexes the amount of information in the subject which is actually being presented by the school. The second indicator is a measure of consistency. Considering as one observation each student's enrollment in each course in the subject, the consistency measure is computed as the product-moment correlation coefficient between the mark received by the students and their overall mark-point averages (MPA's). A high coefficient indicates that students are not performing erratically in the subject. The third indicator is the percent of students who respond that the subject is the one they like best; it indexes interest and successful stimulation in the area. The fourth indicator is the percent of students who indicate that the subject is the one that they like least.

Stimulation Structure. Stimulation Structure clusters consist of the following indicators:

G1-M. Mean Teacher Experience

G1-M.1 Teachers/age: mean of the ages of the teachers

G1-M.2 Teacher experience: mean of the number of years of local experience of the teachers

G1-M.3 Teacher college credit: mean of the number of college credits of the teachers.

G1-V. Variations of Teacher Experience

G1-V.1 Teacher age: variance of the ages of the teachers

G1-V.2 Teacher experience variation: variance of the numbers of years of local experience of the teachers

G1-V.3 Teacher college credit variation: variance of the numbers of hours of college credit of the teachers.

G2-M. Mean Class Heterogeneity

G2-M.1 Class size: mean number of students in the classes

G2-M.2 Class MPA heterogeneity: mean over classes of the variance of the students' MPA's

G2-M.3 Class homework heterogeneity: mean over classes of the variance of the students' amounts of homework, determined according to a six-level response coded from 1 = none to 6 = 21 hours or more

G2-M.4 Class expectations heterogeneity: mean over classes of the variance of the students' educational expectations, determined according to a ten-level response coded from 1 = less than high school completion to 10 = graduate study.

G2-V. Variation of Class Heterogeneity

G2-V.1 Class size variation: variance of the number of students in the classes

G2-V.2 Class MPA heterogeneity variation: variance over classes of the variance of the students' MPA's

G2-V 3 Class homework heterogeneity variation: variance over classes of the variance of the students' amounts of homework, measured as G2-M.3 above

G2-V.4 Class expectation heterogeneity variation: variance over classes of the variance of the students' educational expectations, measured as G2-M.4 above.

The class heterogeneity indicators are based on all classrooms containing at least one student from grades 10-12 inclusive and having an academic subject. For each such class, four measures of heterogeneity were considered: the number of pupils, the variance of the MPA's of the pupils, the variance of the amount of time spent by the pupils on homework, and the variance of the educational expectations of the students. The means over classes of these measures form indicators for cluster G2-M and the variances over classes form indicators for cluster G2-V.

Three measures are used as teacher experience indicators: 1) age; 2) number of years of local teaching experience; and 3) number of college credits. The means and the variances of the measures are defined and included in the mean-level and variation teacher experience clusters. Four measures are defined for each class in a school (classes without any students in grades 10-12 inclusive omitted): 1) number of pupils, 2) the variance of the MPA's of the students, 3) the variance of the amounts of time spent on homework by the students, and 4) the variance of the educational expectations of the students. These are all considered to relate to class heterogeneity; a large class may result in a heterogeneous set of behaviors. The other measures are direct measures of heterogeneity of student characteristics. The mean-level cluster for class heterogeneity has the means over classes of these measures as indicators, and the variation cluster for class heterogeneity has the variances over classes as indicators.

Satisfaction. The satisfaction clusters consist of the following indicators:

H1-M. Mean Attitude Toward Study

H1-M.1 Student attitude: mean response concerning personal like to study, based on a six point scale from 1 = dislike very much to 6 = really like it

H1-M.2 Homework: mean amount of time spent doing homework, based on a six point scale from 1 = none to 6 = 21 hours a week or more

H1-M.3 Parent attitude: mean report of parents' feelings about the amount of homework assigned, based on a four point scale from 1 = indifferent to 4 = too little.

H1-V. Variation of Attitude Toward Study

H1-V.1 Student attitude variation: variance of response concerning liking study, based on a six point scale from 1 = dislike very much to 6 = really like it

H1-V.2 Homework variation: variance of amount of time spent doing homework, based on a six point scale from 1 = none to 6 = 21 hours a week or more

H1-V.3 Parent attitude variation: variance of reports of parents' feelings about amount of homework assigned, based on a four point scale from 1 = indifferent to 4 = too little

H2-M. Mean Social Interaction

H2-M.1 Unpleasant experiences: mean frequency of unpleasant experiences reported with other students, based on a five point scale from 1 = none to 5 = a great many

H2-M.2 Teacher attitude: mean perception of teacher's view of student, based on a five point scale from 1 = very poor to 5 = top

H2-M.3 Activities: mean number of activities in which a student is engaged, based on the number of "very" or "fairly" responses concerning activity in ten extracurricular activities, including athletics, student government, social affairs, etc.

H2-V. Variation of Social Interaction

H2-V.1 Unpleasant experiences variation: variance of the frequency of unpleasant experience reported with other students, based on a five point scale from 1 = none to 5 = a great many

H2-V.2 Teacher attitude variation: mean perception of teacher's view of student, based on a five point scale from 1 = very poor to 5 = top

H2-V-3 Activities variation: variance of number of activities in which a student is engaged, based on the number of "very" or "fairly" responses concerning activity in ten extracurricular activities, including athletics, student government, social affairs, etc.

The attitude-toward-study indicators are based on the responses by the students in grades 10-12 to questions concerning: 1) how much they like to study, 2) how much time each week they spend on homework, and 3) how their parents feel about the amount of homework that is assigned. The first item elicits directly an attitude toward studying. The second is less direct; students who study much usually like to study. The third item was included here as an added indication of student attitude toward studying. Because the items were measures of individuals, their means and variances were computed and assigned to a mean-level and a variation cluster. The social interaction indicators are based on responses by the students to items concerning: 1) how frequent have been unpleasant experiences with other students, 2) what are teacher's atttitudes toward the student, and 3) in what extracurricular activities does the student take part. The first and third measures directly refer to social interaction with other students. The second measure is assumed to be related to the amount of rapport--or social interaction--with teachers. Again, mean-level and variation clusters are defined.

Student Output

The Iowa educational data bank information is insufficient to measure some of the student output clusters, but there was an unexpectedly rich supply of data in the area of cognitive ability and that cluster was expanded. This expansion necessitated a revision of the theoretical scheme for student output. The revised scheme is diagrammed in Figure 5-C-2. The original three ability clusters have been replaced by three cognitive ability clusters. These three clusters correspond to areas covered in the Iowa Test of Educational Development and will not be described further. Also, the aspiration and emotional outlook clusters had to be simplified. One pair of clusters for each replaced the original three pairs for each. The aspiration pair consists of a mean-level and variation cluster for general socioeconomic aspiration relating to educational and occupational goals. The emotional outlook pair is a mean level and variation cluster for general adjustment. Because all these clusters are intended to represent the output of the school system, only twelfth grade students were used in computing the indicators.

Ability			Aspiration			Emotional Outlook		
I1. Verbal Ability	M	V	J1. Aspiration	M	V	K1. Adjustment	M	V
I2. Science and Quantification Ability	M	V						
I3. Social Science Ability	M	V						

Figure 5-3-2

Revised Clusters for Student Output

Ability. The ability clusters consist of the following indicators:

I1-M. Mean Verbal Ability

- I1-M.1 Expression: mean score on ITED-3, correctness and appropriateness of expression
- I1-M.2 Vocabulary: mean score on ITED-8, general vocabulary
- I1-M.3 Literature: mean score on ITED-7, literature interpretation

I1-V. Variation of Verbal Ability

- I1-V.1 Expression variation: variance of scores on ITED-3, correctness and appropriateness of expression
- I1-V.2 Vocabulary variation: variance of scores on ITED-8, general vocabulary
- I1-V.3 Literature variation: variance of scores on ITED-7, literature interpretation

I2-M. Mean Science and Quantification Ability

- I2-M.1 Science Background: mean score on ITED-2, general background in the natural sciences
- I2-M.2 Quantification: mean score on ITED-4, ability to do quantitative thinking
- I2-M.3 Science Interpretation: mean score on ITED-6, interpretation in natural science

I2-V. Variation of Science and Quantification Ability

- I2-V.1 Science background variation: variance in scores on ITED-2, general background in the natural sciences
- I2-V.2 Quantification background variation: variance of scores on ITED-4, ability to do quantitative thinking
- I2-V.3 Science interpretation variation: variance in scores on ITED-6, interpretation in natural science.

I3-M. Mean Social Science Ability

- I3-M.1 Social concepts: mean score on ITED-1, understanding of basic social concepts
- I3-M.2 Social studies interpretation: mean score on ITED-5, interpretation in social studies
- I3-M.3 Information source utilization: mean score on ITED-9, use of sources of information.

I3-V. Variation of Social Science Ability

I3-V.1 Social concepts variation: variance of scores on ITED-1, understanding of basic social concepts

I3-V.2 Social studies variation: variance of scores on ITED-5, interpretation in social studies

I3-V-3 Information source utilization variation: variance in scores on ITED-9, use of sources of information.

The nine subtests on the Iowa Test of Educational Development were partitioned into three groups: 1) verbal ability, 2) science and quantification ability, and 3) social science ability. Each group consists of three tests. For each group, a pair of clusters is defined; one cluster for means and one for variances.

Aspiration. The aspiration clusters consist of the following indicators:

J1-M. Mean Aspiration

J1-M.1 Educational expectation: mean education level the students expect to attain, based on a 10 point scale from 1 = less than high school graduation to 10 = graduate study

J1-M.2 Educational discrepancy: mean discrepancy between the education levels expected by the students and those achieved by their fathers

J1-M.3 Girls' aspirations: mean occupational aspiration of girls, based on a five point scale from 1 = full-time homemaker to 5 = full-time career.

J1-V. Variation in Aspiration

J1-V.1 Educational expectation variation: variance of the education levels the students expect to attain, based on a 10 point scale from 1 = less than high school graduation to 10 = graduate study

J1-V.2 Educational discrepancy variation: variance of the discrepancies between the educational levels expected by the students and those achieved by their fathers

J1-V.3 Girls' aspirations variation: variance of occupational aspirations of girls, based on a five point scale from 1 = full-time homemaker to 5 = full-time career.

For each student in twelfth grade, three measures are computed: 1) educational expectation level, 2) discrepancies between expected educational level and educational level of the student's father, and 3) occupation aspiration of girls. The educational expectation level generally indicates the socioeconomic aspiration of the student. The discrepancy with the father's educational level refines this by indicating the degree to which the expectation is upward-mobile. Occupational aspiration information was not available for boys, but for girls an item response indicated whether the girl aspired to a career, full-time or part-time, or to be a homemaker. The career aspiration of girls may be considered an indicator of the general socioeconomic aspiration of the students. Because these measures are individualistic, two clusters were formed, one for mean-level and one for variation.

Emotional Outlook. The emotional outlook clusters consists of the following indicators:

K1-M. Mean Adjustment

- K1-M.1 Markpoint variance: mean student markpoint variance, computed as the variance of the marks that a student receives in academic courses
- K1-M.2 Aspirational discrepancy: mean difference between what educational levels students believe they could attain and what levels they expect to attain, based on a 10 point scale from 1 = less than high school graduation to 10 = graduate study
- K1-M.3 Uncertainty: percent of students who are uncertain of their post-high school educational plans
- K1-M.4 Capacity learning: correlation coefficient between MPA's and composite ITED scores

K1-V. Variation in Adjustment

- K1-V.1 Markpoint variance variation: variance of student markpoint variances, computed as the variance of the marks that a student receives in academic courses
- K1-V.2 Aspirational discrepancy variation: variances of differences between the educational levels that students believe they could attain and what levels they expect to attain, based on a 10 point scale from 1 = less than high school graduation to 10 = graduate study.

For each student in the twelfth grade, a special mark summary coefficient was formed--the markpoint variance. This was computed as the variance of the marks received by the student in academic courses, and it is considered to be a measure of consistency of performance. It is assumed that consistency is an aspect and indicator of adjustment. Another indicator was computed from two parallel questions asked of the students: what educational level they expected to attain, and what level they thought they could attain. The discrepancy between the responses is also considered a measure of adjustment. The means and variances of these two measures were computed over students and assigned respectively to the mean-level and variation adjustment clusters. Two additional measures were computed to indicate mean-level of adjustment. The first is a measure of uncertainty; i.e., percent of students who were uncertain of their future educational plans. The second is a measure of the discrepancy between test score and MPA: specifically, it is the correlation coefficient between the two. When there is a high correlation, then it can be said that students are working to capacity. A lower coefficient may indicate emotional distraction or interference.

D. Analytic Techniques

This section contains the theoretical bases, statistical properties, computational formulas, and interpretative implications of the analyses performed on the data. The indicator variables were extracted from the Iowa and U.S. Census data banks by a series of computer programs, which are briefly described below, and were fed through a four-stage computing procedure before the final results were obtained. The first stage--filtering--has not been discussed previously, since its need became apparent only during the final stage of the project. The second stage involved compositing the indicators by clusters according to principal component analysis. The third stage involved factoring the clusters within each supercluster--Community Input, School Mediation, and Student Output--according to image analysis. The fourth stage involved regression and graphical analysis of the interrelationships of the image factors.

The principal component compositing, the image analyses, and some of the stage four analyses were first performed on indicator variables as calculated from essentially the entire population of Iowa secondary schools. It was apparent that there were complicating data in the indicator data matrix. In particular, image factor scores, which automatically have means of zero and standard deviations of one, were found for some schools to exceed plus or minus ten. This indicated enormously skewed distributions, even though after the compositing operation one would have expected approximately normal distributions. There were two possible reasons for the skewness: either there were schools which were most unusual in terms of their educational situations or qualities, or there were erroneous data mixed in with good data. Consequently, a manual examination was made of the data for schools with extreme factor scores. (Careful monitoring of all the input data was not possible because of the vast amount of data processed and because of the desire to establish the probability of just these kinds of problems.) It was found that the extreme values generally corresponded to missing or inconsistently coded data and not to unusual data. Some procedure had to be developed for omitting such data or for providing suitable substitutes.

Except in a few isolated cases, in which census data were incorrect because of manual errors in the mapping operation, all the errors were found in the Iowa educational data bank data. It should be stated that the errors encountered are not those which would significantly affect the utility or value of the Iowa educational data bank. In fact, the personnel at the data bank operate with the expectation that data will occasionally be missing or incorrect. The amount of time spent correcting data is proportional to the necessity that the data be correct. For example, it is necessary that each teacher's social security number be recorded correctly. If, however, a teacher's current assignment is incorrect or missing, no irreparable damage is done. The objective of data processing systems is service. If a school fails to convert its grading codes to the statewide code, the most that can happen in terms of service is that the school will not receive computed grade points. The data received from the Iowa bank is certainly sufficiently accurate and complete for administrative purposes. A question must be raised as to whether it is satisfactory for research problems such as the one reported herein.

Two problems were envisioned had the erroneous data been retained and ignored. First, they generally depress correlations. Since from preliminary analysis it was known that the correlations would be low, this was not acceptable. Second, graphical presentations were planned. Therefore, such data were not acceptable because the scaling of the graphs would have been overextended. Third, such data make inappropriate the definition and computation of orthogonal factor scores. This last point is not widely known, but may be illustrated by an example. Given two orthogonal components derived from a set of variables by the method of principal components, one might expect the variates to be bivariate normally distributed, with correlation zero. Yet, all orthogonalization guarantees is that the linear correlation is zero, while the bivariate distributions may consist of the mixture of a bivariate normal distribution, with high correlation, and one extreme outlier placed in such a way as to negate the correlation of the other points.

The problems envisioned were considered serious enough so that extreme measures were employed to clean the data. The operation is called filtering and is described later in this section. It consisted of omitting all schools which were in the extreme tails of the distributions for one or more variables. This operation might not have been necessary if different kinds of analysis had been employed. In this research, the school is the unit under study and the sample size is actually quite small, being in the hundreds. If students were the units of study, such data would have been less serious, since students numbered in the hundred thousands. The particular analytic techniques used here were especially sensitive to erroneous data. Yet, it must be concluded that the data should probably be considered inadequate for the purposes of the present research. By the end of the filtering operation, about half the schools had been omitted from the analysis. The investigators cannot answer the crucially important question of whether schools having such data correlate with the high school size/non-dollar costs relationship; i.e., do the schools omitted from consideration form a special sub-population of schools in terms of their educational or environmental characteristics? Does failure to supply data indicate special qualities which are related to how schools are operated and, in turn, to how students are taught and learn?

Construction of the Indicators

Prior to performing any of the analyses, it was necessary to run a series of computer programs which extracted the indicator information from the data bank tapes. In the following paragraphs are given brief descriptions of the programs and their purposes. All the programs were written in a combination Fortran-63 and Compass machine language for the CDC 3600.

Census data processing. Three programs ran sequentially to produce the indicators use in Community Input. The first program read the census data tapes and the mapping cards. The contents of the mapping cards are described in Chapter 4; essentially, there was one mapping card for each census unit, giving the proportional assignment of the unit to school areas. The first program collated the two files. It produced an output tape file which consisted of a copy of the census data plus tags indicating for each unit which schools with which proportions were assigned to the unit. The second program then made repeated passes through the new tape file. There was one pass for each successive group of 28 schools--this limit was imposed by computer memory size. The values of the census items for the 28 schools were accumulated internally during a pass. As each tagged data set was read, its tags were examined to see whether or not the unit was assigned, to any extent, to any of the 28 schools; if it was, its values were added to the accumulated sums for the school after multiplying by the indicated proportions. The output tape file of the second program served as input to the third program, which formed the ratios, percentages, etc., needed as the Community Input indicators.

Student data processing. The student data were recorded on a 13-tape file. One computer program was used to process the information and to produce most of the indicators for School Mediation and Student Output. The program was run 13 times, once for each tape in the file. The program read the student records school by school. For each school it determined the averages, proportions and variances required for construction of indicators. In addition, it reconstructed the assignment of students to classrooms on the basis of student course records, and determined the averages and variances of classroom qualities. For each school, the program generated the constructed indicators.

Teacher data processing. Two clusters of indicators in the School Mediation group required examinations of the Teacher data file. For this purpose a program was run which read the 2-tape teacher file and computed the necessary means and variances by school.

Filtering the Data

The data processing operations described above yielded a data matrix on tape of about 120 indicator variables and about 400 schools. As noted above, there were many missing or miscoded data. A quick visual examination was first made of a printout of the data. This led to the omission of about 70 schools. A preliminary set of principal component and image analyses was then made. At that time, the extreme factor scores were detected during a second visual examination of a printout. A detailed examination was made of the data for which the largest extreme scores had been obtained. It was found that, in all cases, those schools had consistently miscoded or missing data. In one school, no seniors had filled out the general information questionnaire. In another school, only half the students had grades reported. The large amount of missing and incorrect data in these schools suggested that scores other than the extreme ones noted might be spurious. The only possible solution was to omit those schools.

Because of the amount, it was not possible to check all the source data manually. The existence of undetected erroneous data was considered sufficiently certain that it was necessary to devise an automatic procedure for clearing the data matrix of schools with such data. The "filtering" operation consisted of a computer program which scanned the indicator data matrix and computed the means and standard deviations of each indicator variable. A second pass was then made through the matrix. This time, each variable for each school was compared to the mean and standard deviation of the variable for the population. If one or more of the values exceeded the mean by 3.5 or more standard deviations, the school was omitted. In all, 110 schools were omitted, and the final sample size for analysis consisted of 220 schools.

The 3.5 standard deviation criterion would correspond roughly to the omission of 5% of the schools in the case where 120 independent normally-distributed variables are processed. Because of the nonnormality and dependency of the data, roughly 30% of the schools were omitted. This must be considered somewhat detrimental in terms of the representativeness of the sample of schools. The only apparent alternative was to have transformed the distributions of the variables; i.e., to have substituted standard values of 3.5 for standard values of more than 3.5. However, because of the correlated nature of the questionable data--e.g., if a school was missing a large number of questionnaire responses, many variables would be affected--substitution would probably have led to less obvious but just as spurious final results.

Composition of the Clusters

Given a set of variables, the principal component, as defined by Hotelling (1935), is a certain linear combination of the standardized versions of the variables. The weights in the combination are determined by the largest eigenvalue of the correlation matrix of the variables and by its corresponding unitized eigenvector. Specifically, the weight for the ith standardized variable is equal to the ith entry in the eigenvector divided by the square root of the eigenvalue. The correlations between the variables and the component score are similarly definable. The correlation between the ith item and the score is equal to the ith entry in the eigenvector multiplied by the square root of the eigenvalue.

The principal component is the linear combination of the variables which represents the maximum amount of variance that can be extracted, by a single linear combination, from the variables. This can be understood best in the sense of correlation between the variables and the component score. With the principal component, the sum of the squares of these correlations is maximized. When the eigenvalue on which the principal component is based is compared to the number of variables, the result is the proportion of the total variance of the variables which is extracted by the principal component.

When the indicators of a cluster have homogeneously high correlations, the proportion of variance extracted by a principal component analysis is high. For example, for the (hypothetical) correlation matrix

1.0	.8	.8
.8	1.0	.8
.8	.8	1.0

82.5% of the variance is extracted. If the correlations are homogeneous, but low, the principal component is still an efficient "blend" of the variables, though a smaller proportion of the variance is extracted. There are problems, however, if the correlations are heterogeneous. Two cases should be considered. First, there may be distinct structures in the correlation matrix, and the principal component may extract and represent only one of the structures. For example, the principal component of this matrix

1.0	.8	.0	.0
.8	1.0	.0	.0
.0	.0	1.0	.6
.0	.0	.6	1.0

correlates only with the first two variables. Second, there may be systematically small correlations but with different signs, and the principal component may correlate positively with some variables and negatively with others. For example, for this correlation matrix

1.00	.06	-.10
.06	1.00	-.08
-.10	-.08	1.00

the principal component correlates positively with the first two variables, but negatively with the last. This would be considered undesirable if the three variables were to be considered positive but independent aspects of some quality--a hypothesis not disproved by the signs because of the small correlations.

Many instances of these effects are realized in the data analysis for this project, yet principal components analysis was the only available appropriate analysis. There is need, therefore, to promote the development of a new statistical technique. The problem is to devise an automatic weighting procedure for compositing heterogeneously correlated variables--for determining a score which correlates highly and with specified directions with all the variables in the set. One criterion which might be used may be called "maximin". The criterion is that all variables should have the specified sign and the smallest correlation (in absolute value) should be maximized. A computational procedure for the maximin component has not been determined, although considerable time has been expended in trying to find one and in trying to identify the properties the maximin component would have. It is not clear whether a properly signed component always exists or whether the maximin is unique. If, however, such a component technique can be found, it would be a useful addition to the indicator-cluster-factor methodology employed in this project. Perhaps for each cluster, the decision to use principal or maximin component analysis would be made on the basis of the theory for clustering the indicators or on the basis of the determined correlation matrix.

However, principal component analyses were used on the data for this project, and it will now be described how such analyses can be interpreted. First, the proportion of variance extracted is a measure of the homogeneity and magnitude of the correlations. A high proportion indicates that the principal component score is a good surrogate for the set of variables. In terms of predicting some outside variable, the principal component will probably serve almost as well as the entire set of variables. On the other hand, a low proportion indicates that a considered amount of the predictive power of the indicators has been lost in the compositing. The correlations or loadings between the component and the variables indicate the degree to which the individual variables are included in the composite and the directional relationship that they have with the component. Certain variables may have negative loadings and others may have essentially zero loadings. The pattern of correlations is examined for the purpose of reinterpreting the cluster; if the signs or magnitude of the loadings are not as had been expected, the meaning of the cluster as measured may not be the same as the meaning originally intended. The name of the cluster may then have to be redefined.

Image Analysis

The term "factor analysis" refers to two related models: "pure" factor analysis, in which there is one unique factor for each variable as well as a set of common factors, and "component analysis", in which there are only common factors and they are combinations of the variables. With pure factor analysis (alpha and canonical factor analysis) a relatively reasonable model is achieved, but factor scores are not computable. With component analysis, a less reasonable model is achieved, but factor scores are computable. Because the factor scores were to be used in the stage four computations of the project, it was necessary to use a component analysis model. The component analysis approach chosen was image analysis, which was originally formulated by Guttman (1953) and later extended and improved by Harris (1962). Image analysis is a component analysis technique which is intended to approximate pure factor analysis.

In image analysis, there is an implicit transformation of the variables. Each variable is transformed implicitly into its "image", which is defined as its regression estimate--i.e., the variable as predicted from the other variables. Also defined implicitly is the "anti-image" which is the residual in the regression. In image analysis, the factors are determined on the basis of the images rather than on the basis of the original variables. Because of certain mathematical identities, it is not necessary explicitly to compute the image variables. All the computations can be carried out in terms of the original variables.

Let Z be the p x N matrix of standard (means = 0, variances = 1) original variables. The correlation matrix is $R = N^{-1}ZZ'$. From the correlation matrix, the image analysis may proceed. First, the diagonal matrix of anti-image variances is computed: S^2 = the diagonal matrix of the reciprocals of the diagonal entries of R^{-1}. Note that a diagonal entry of S^2 is the complement of the squared-multiple-correlation between the corresponding variable and all the other variables. It would be possible at this point to compute and factor the image covariance matrix: $G = R + S^2R^{-1}S^2 - 2S^2$. However, according to the Harris scheme, a different matrix is factored: $S^{-1}RS^{-1} = XB_r^2X'$, where B_r^2 is the diagonal matrix of eigenvalues of $S^{-1}RS^{-1}$ and the columns of X are the corresponding unitized eigenvectors. Let $B_g^2 = (B_r^2 - I)^2B_r^{-2}$. Then the Harris factors of G are $F_g = SXB_g$. Actually only certain of the factors are retained. All factors corresponding to roots in B_r^2 greater than one are retained. These factors may then be rotated by Kaiser's (1958) normal varimax procedure and F_gT is obtained. It can be shown that the image factor scores can be computed from the original variables according to this formula: $Y = T'B_rX'S^{-1}Z$. These factor scores have means equal to zero, variances equal to one, and intercorrelations equal to zero.

Guttman (1953) originally showed that image analysis is an approximation of pure factor analysis. When a universe of variables is sufficiently sampled, the image factors approximate the common factors. Guttman further showed that this approximation is appropriate when the correlations between the anti-images are close to zero. The anti-image correlation matrix is $SR^{-1}S$. Note that $SR^{-1}S$ is the inverse of the matrix, $S^{-1}RS^{-1}$, which was factored. This implies that the small roots of $S^{-1}RS^{-1}$ correspond to the large roots of $SR^{-1}S$. By omitting the small roots of $S^{-1}RS^{-1}$, the large roots of SR^{-1} are omitted from the data analysis model; i.e., the reproduced anti-image correlation matrix would have, as Guttman shows to be important, small off-diagonal entries. This approach was developed by Harris (1962) and he also provides further justification for it.

In Section 6-3, four matrices are presented for each image analysis. First, there is the correlation matrix of the original variables. Second, the diagonal entries of S^2 (i.e., the anti-image variances or the complements of the squared-multiple-correlations) are given. Third, there are the Harris roots of G: B_g^2. Finally, the rotated image factor matrix, F_gT, is given. The anti-image variances indicate the degrees to which the original variables are not predictable from one another--i.e., the proportions of the variables which approximate unique variance. The magnitudes of the omitted roots in B_g^2 (roots corresponding to roots in B_r^2 less than one are omitted) indicate the amount of anti-image correlation. The higher these values, the less valid is the approximation to pure factor analysis and the less likely it is that the universe of variables has been properly sampled. The factor matrix gives the cross-covariances of the image variables and the rotated factors. The meaning of a factor is determined on the basis of the variables which have the highest covariances or loadings. The factors which are most important are those which have the highest sums of squares of loadings. The variables which are essentially ignored in the analysis are those which have the lowest sums of squares of loadings.

Regression Analysis

Although planned, no extensive advanced regression analysis was used. To get a sense of the data, a set of simple multiple regressions was performed. These techniques are well-known and require no discussion.

Contour-Plotting

Somewhat tangentially to the regular course of this research, a statistical technique was developed and tested. The technique is not new but rather is an application of a computer technique often used in geophysical research. A geophysicist might have gravity readings over an area and want to diagram the hypothetical gravity surface to aid his search for anomalies. To provide such diagrams, geophysicist have developed computer programs which accept as input levels each point on a grid and which produce as output a contour map of the surface. A contour map is a map in which lines are drawn connecting places of equal level or value.

The statistical application of contour plotting concerns the examination of the relation of two variables as in diagramming their joint distribution. One solution commonly employed is to produce a scatter plot. However, this is not generally appropriate when there are many observations. Even when there are few, scatter plots only suggest the term of a distribution. However, if the data are plotted on a grid and each variable is transformed by a linear function into the range 1-n where n is some chosen integer, a contingency table may be produced. A contour plot may then be made, considering the grid points as equally spaced and the frequencies as levels. The data are then in the form acceptable to a geophysical plotting program and the shape of the distribution can be explicitly diagrammed.

This technique seems to have great promise in terms of finding and defining non-linear distributions. A more important application may be in detecting mixture distributions--i.e., distributions which are the sum of several different distributions of several different subpopulations. Two examples of frequency distribution contour plots are given in Section 6-C.

CHAPTER 6

RESULTS AND INTERPRETATIONS

This chapter contains four sections. In the first section, the principal components analysis which was performed on the clusters of indicators is described. The second section describes the factor structures. The third section contains the regression structures. The fourth section is a concluding statement.

A. Properties of the Clusters

In this section are tabulated the results of the principal component analyses performed on the clusters of indicators. The discussion is divided into three parts--one for Community Input, one for School Mediation, and one for Student Output. Principal component analysis seems to be inappropriate for clusters which contained non-homogeneously correlated indicators and the principal component fails to correlate with or correlates with the wrong sign of some of the indicators. The discussion is aimed, therefore, at redefining or clarifying the meaning of the cluster measurements in terms of their actual computational relationships with the indicators. For each group of clusters, the means, standard deviations, intercorrelations, and principal component loadings of the indicators are given for each cluster within the group.

Community Input

The Community Input clusters were more appropriately analyzed by principal component method than the clusters of the other superclusters, since higher and more homogeneous inter-indicator correlations were found. The analytic results are discussed in the following paragraphs.

Population Dynamics. The analytic results for these clusters are presented in Table 6-A-1. The three growth indicators have a non-transitive correlation structure: birth rate and immigration are each correlated with residence newness, but they are not correlated with each other. This is reasonable since a high birth rate and a high amount of immigration are each an independent source of residence construction. The principal component loads highly on all three indicators. The structure of the youth indicators is more complicated. Infants and children are correlated as are young men and young women; also, infants is correlated with young men and young women, but children

TABLE 6-A-1

CLUSTER ANALYSIS RESULTS FOR
COMMUNITY-INPUT--POPULATION DYNAMICS

Clusters		Mean	S. D.	Correlations				Principal Component
A1. Growth (48.3% of the variance was extracted)								
A1.1	Birth Rate	21.50	3.48	---	.35	-.05		.62
A1.2	Residence Newness	11.27	5.37	.35	---	.32		
A1.3	In-migration	13.43	4.12	-.05	.32	---		
A2. Youth (53.2% of the variance was extracted)								
A2.1	Infants	12.99	1.49	---	.51	.44	.37	.78
A2.2	Children	26.15	1.89	.51	---	.13	-.05	.42
A2.3	Young Males	4.58	.66	.44	.13	---	.72	.85
A2.4	Young Females	4.93	.67	.37	-.05	.72	---	.78
A3. Density								
A3.1	Structure Density	1.11	.15	---	-.40	.15		-.43
A3.2	Unit Density	3.26	.23	-.40	---	.55		.94
A3.3	Room Density	.49	.03	.15	.55	---		.74

as an indicator is not. This complication is doubtless related to the presence of families with infants yielding high proportions of young adults and infants and to a general tendency for a community to have many or few infants and children. All four indicators have high correlation with the principal component. The density cluster exhibits a non-transitive correlation structure: structure density and room density are each correlated with unit density but not with each other. It should also be noted that the correlation between structure density and unit density is negative. Unit density (persons per unit) is high when structure density (units per structure) is low or when room density (persons per room) is high. These empirical results imply that density should be interpreted in terms of many people in a small living area.

Population Characteristics. The analytic results for these clusters are presented in Table 6-A-2. The rural/urban cluster measurement was computed in such way that rural areas have high scores. The correlational structure of the three indicators is non-transitive. Urban residence is negatively correlated with rural non-farm residence and agriculture employment, but the latter two are essentially uncorrelated. This result suggests that in a non-urban area, non-farm residence and employment in agriculture are essentially independent. The principal component correlates highly with each of the indicators, the correlation being negative for urban residence. The socioeconomic level cluster is somewhat confused because of the difficulty of coding occupation levels in rural areas. The farming occupation was given a relatively high occupation score, yet farmers generally have less income and education than high-status urban workers. In this cluster then, occupation level correlates negatively with income level. However, the principal component correlates positively and highly with income level, and is probably an adequate measure of socioeconomic status. The indicator correlations for socioeconomic variation are essentially zero. The pattern of signs is the same as for the previous cluster, and the principal component may be considered to represent variation in socioeconomic standing.

Socialization Characteristics. The analytic results for these clusters are presented in Table 6-A-3. For the social activity involvement cluster, the social workers indicator is uncorrelated with the other indicators and has a low correlation with the principal component. On the other hand, public employees and female employees are correlated, and both correlate highly with the principal component. However, the meaning of this cluster is obscure. The two indicators of the social mobility clusters are essentially uncorrelated, but what correlation exists is negative. This makes the indicators correlate with the principal component in opposite directions; i.e. an area with a high score on the principal component would probably have a low heterogeneity but a high population mobility. This must be related to the urbanization processes in Iowa. There are three poverty culture indicators, and the last, unemployment, is essentially uncorrelated with the other two and has a low correlation with the principal component. Apparently, unemployment is not a regular problem in Iowa. The other two indicators, low income, and poor housing are correlated and both have high correlations with the principal component.

TABLE 6-A-2

CLUSTER ANALYSIS RESULTS FOR
COMMUNITY-INPUT--POPULATION CHARACTERISTICS

Clusters	Mean	S. D.	Correlations			Principal Component
B1. Rural/Urban (69.9% of the variance was extracted)						
B1.1 Urban Residence	13.70	26.32	---	-.80	-.67	-.99
B1.2 Rural Non-Farm Residence	40.89	19.13	-.80	---	.12	.79
B1.3 Agriculture Employment	40.69	14.60	-.67	.12	---	.69
B2. Socioeconomic Level (58.8% of the variance was extracted)						
B2.1 Income Level	4.99	.55	---	.39	-.60	.91
B2.2 Education Level	5.02	.20	.39	---	-.10	.56
B2.3 Occupation Level	7.40	.48	-.60	-.10	---	-.79
B3. Socioeconomic Variation (38.9% of the variance was extracted)						
B3.1 Income Variation	8.12	1.01	---	.06	-.10	.63
B3.2 Education Variation	2.02	.21	.06	---	-.08	.57
B3.3 Occupation Variation	9.20	1.11	-.10	-.08	---	-.67

TABLE 6-A-3

CLUSTER ANALYSIS RESULTS FOR
COMMUNITY INPUT--SOCIALIZATION CHARACTERISTICS

Clusters	Mean	S. D.	Correlations			Principal Component
C1. Social Activity/Involvement (45.5% of the variance was extracted)						
C1.1 Social Workers	2.25	1.69	---	.13	.00	.30
C1.2 Public Employees	2.56	1.40	.13	---	.34	.83
C1.3 Female Employees	35.74	7.84	.00	.34	---	.77
C2. Social Mobility (57.2% of the variance was extracted)						
C2.1 Population Heterogeneity	15.73	6.98	---	-.14		-.76
C2.2 Population Mobility	22.93	4.69	-.14	---		.76
C3. Poverty Culture (45.2% of the variance was extracted)						
C3.1 Low Income	22.10	6.25	---	.32	-.03	.72
C3.2 Poor Housing	41.62	11.13	.32	---	.19	.84
C3.3 Unemployment	2.68	1.73	-.03	.19	---	.37

Economic Structure. The analytic results for these clusters are presented in Table 6-A-4. In the wealth cluster, property value and rent value have a substantial intercorrelation and are both highly correlated with the principal component. But automobile availability has very low correlations with property value and rent value and only a moderate correlation with the principal component. Apparently, the availability of automobiles is confounded by other factors--perhaps the age and type of automobile. Both the industry and trade clusters consist of highly and homogeneously correlating indicators, and the correlations with the principal components are high. The distinction made between industry and trade is not clear. The farming cluster consists of variables with low intercorrelations. However, the principal component seems logical. It correlates negatively with family farm help and positively with paid farm help and food industry. It would seem, then, that an area with a high value on the component would have a commercial farming industry rather than family farming.

Educational Resources. The analytic results for these clusters are presented in Table 6-A-5. The first cluster, archive usage, contains just one indicator, librarians, and the principal component is equal to the indicator. However, the indicator librarians is probably without value, since the census data on which it is based are too gross. The school areas are relatively small, the count of employed librarians is based on a 1-in-4 sample, and librarians are rare. Therefore, the estimated number of librarians per 1,000 population tends to be inaccurate. The two educational involvement indicators, school enrollment and school employment are essentially uncorrelated. This is probably due to the fact that rural teachers do not always live in the area where they teach. Both indicators correlate highly with the principal component. The two indicators for the educational attainment cluster, college attainment and professional employment, have a high intercorrelation and both correlate highly with the principal component.

School Mediation

It must be admitted that the measurement achieved of the school mediation clusters was generally imperfect. Part of the reason was the absence, noted earlier, of data appropriate to measure the kinds of social processes theoretically considered important. Part of the reason was the lack, also noted earlier, of an analytic technique for compositing non-homogeneously correlating variables. Another basic problem with the data bank information became apparent upon examination of the cluster analysis results and the associated source data. All but the most generalized curricular control is exercised by local school officials. Consequently, course contents, requirements, and standards vary widely. On the other hand, in fitting the hundreds of curriculum patterns into a statewide code, many quite inappropriate equivalences had to be assumed by the founders of the data bank. For example, English IV may be a course required of everyone in one district and a course taken only by the best five percent of students in another. The general problem is that

TABLE 6-A-4

CLUSTER ANALYSIS RESULTS FOR
COMMUNITY INPUT--ECONOMIC STRUCTURE

Clusters	Mean	S. D.	Correlations			Principal Component
D1. Wealth (50.8% of the variance was extracted)						
D1.1 Property Value	2.84	.73	---	.45	.12	.81
D1.2 Rent Value	6.02	.78	.45	---	.15	.83
D1.3 Automobile Availability	1.12	.10	.12	.15	---	.43
D2. Industry (65.2% of the variance was extracted)						
D2.1 Industry Employment	22.82	7.95	---	.34	.63	.83
D2.2 Managerial Employment	6.34	2.43	.34	---	.44	.70
D2.3 Clerical Employment	7.53	2.95	.63	.44	---	.88
D3. Trade (75.3% of the variance was extracted)						
D3.1 Retail Employment	14.45	4.67	---	.58	.73	.89
D3.2 Service Employment	5.52	2.08	.58	---	.58	.82
D3.3 Sales Employment	5.36	2.35	.73	.58	---	.89
D4. Farming (40.7% of the variance was extracted)						
D4.1 Family Farm Help	.11	.07	---	-.19	-.02	-.69
D4.2 Paid Farm Help	.18	.08	-.19	---	.10	.76
D4.3 Food Industry	3.44	2.69	-.02	.10	---	.40

TABLE 6-A-5

CLUSTER ANALYSIS RESULTS FOR
COMMUNITY INPUT--EDUCATIONAL RESOURCES

Clusters	Mean	S. D.	Correlations		Principal Component
E1. Archive Usage (100% of the variance was extracted)					
E1.1 Librarians	.32	.59	---		1.00
E2. Educational Involvement (53.5% of the variance was extracted)					
E2.1 School Involvement	25.73	2.83	---	.07	.73
E2.2 School Employment	5.48	1.91	.07	---	.73
E3. Educational Attainment (74.5% of the variance was extracted)					
E3.1 College Attainment	13.15	3.56	---	.49	.86
E3.2 Professional Employment	7.01	2.30	.49	---	.86

although the data bank codings have administrative meaning--e.g., in both cases above, "English IV" is the proper printout for class lists--they do not always have substantive meaning.

Stimulation Substance. The analytic results for these clusters are given in Table 6-A-6. The problem of general coding inconsistency discussed above is especially important for these clusters. It was originally intended that these clusters represent different substances and not be tied to administrative distinctions such as courses. For example, literature discussed in a history course was English stimulation nonetheless. Since this was not possible, the administrative distinctions had to be used. Even then, the curricular differences between schools has essentially prevented any measurement of stimulation substance. The indicators within each of the five clusters have very low correlations. It does not seem profitable to attempt explanation of the patterns of correlations or to interpret the meaning of the principal components. Although these concepts are doubtless essential to the question of high school size and non-dollar costs--e.g., large schools may very well provide more stimulation in certain areas--they could not be measured.

Stimulation Structure. The analytic results for these clusters are presented in Table 6-A-7. The correlations for the mean and variation teacher experience clusters are fairly high and all the indicators correlate highly with their respective principal components. The indicators selected for classroom hererogeneity seem to be essentially uncorrelated both for mean-level and variation. As noted, principal components analysis yields spurious results in such cases; only 33 percent and 27 percent of the variances of the indicators is extracted.

Satisfaction. The analytic results for these clusters are presented in Table 6-A-8. For the first cluster, mean attitude toward study, an indicator was omitted and the resulting two indicators do not define a meaningful concept. The second cluster, variation in attitude toward study, has low interindicator correlations, and principal components analysis has yielded a component with uninterpretable signs. The mean social interaction cluster has one moderate interindicator correlation between teacher attitude and activities. There is apparently a more comfortable social relationship between teachers and students when they interact in activities. Both indicators correlate highly with the principal component. The indicators for variation of social interaction have positive correlations and correlate highly with the principal component. The composite measure apparently indexed variation in social attitude and participation--i.e., the presence of both activists and loners.

Student Output

For student output, the more concrete clusters, which involved an established test battery and a definite and obvious aspiration scale, seemed quite amenable to the principal component analysis. The abstract clusters, which represented attempts indirectly to measure emotional qualities, yielded equivocal results.

TABLE 6-A-6

CLUSTER ANALYSIS RESULTS FOR
SCHOOL MEDIATION--STIMULATION SUBSTANCE

Cluster	Mean	S. D.	Correlations				Principal Component
F1. Mathematics (32.[illegible] of the variance was extracted)							
F1.1 Advanced Mathematics	21.07	10.84	---	-.03	.09	-.09	.45
F1.2 Math. Consistency	.77	.10	-.03	---	.04	.05	-.06
F1.3 Mathematics/Like	14.73	4.86	.09	.04	---	-.26	.74
F1.4 Mathematics/Dislike	24.64	6.79	-.09	.05	-.26	1.00	-.75
F2. Science (34.6% of the variance was extracted)							
F2.1 Advanced Science	4.20	8.19	---	-.08	-.09	.03	.23
F2.2 Science Consistency	.81	.09	-.08	---	.00	.08	.11
F2.3 Science/Like	12.38	4.95	-.09	.00	---	-.36	-.81
F2.4 Science/Dislike	13.11	5.40	.03	.08	-.36	---	.81
F3. English (38.4% of the variance was extracted)							
F3.1 Advanced English	29.85	16.47	---	-.11	.12	-.13	.44
F3.2 English Consistency	.81	.09	-.11	---	-.20	.08	-.49
F3.3 English/Like	11.77	4.46	.12	-.20	---	-.37	.76
F3.4 English/Dislike	18.21	6.61	-.13	.08	-.37	---	-.71
F4. History (32.6% of the variance was extracted)							
F4.1 Advanced History	32.44	13.08	---	.07	-.05	.13	.50
F4.2 History Consistency	.81	.07	.07	---	.04	.15	.45
F4.3 History/Like	10.23	4.31	-.05	.04	---	-.20	-.51
F4.4 History/Dislike	18.73	6.73	.13	.15	-.20	---	.76
F5. Foreign Language (37.5% of the variance was extracted)							
F5.1 Foreign Language	13.55	7.69	---	.33	.32	.04	.82
F5.2 For. Lang. Consistency	.71	.30	.33	---	.05	.15	.67
F5.3 For. Lang./Like	2.09	1.95	.32	.05	---	-.03	.57
F5.4 For. Lang./Dislike	9.48	4.85	.04	.15	-.03	---	.23

TABLE 6-A-7

CLUSTER ANALYSIS RESULTS FOR
SCHOOL MEDIATION--STIMULATION STRUCTURE

Cluster	Mean	S. D.	Correlations				Principal Component
G1(M). Mean Teacher Experience (59.9% of the variance was extracted)							
G1(M).1 Teacher Age	37.74	3.86	---	.47	.31		.77
G1(M).2 Teacher Experience	4.25	1.76	.47	---	.41		.83
G1(M).2 Teacher College Credit	152.57	7.72	.31	.41	---		.72
G1(V). Variation of Teacher Exper. (49.5% of the variance was extracted)							
G1(V).1 Teacher Age Var.	165.04	90.02	---	.35	.28		.82
G1(V).2 Teacher Experience Variation	30.89	26.41	.35	---	.08		.68
G1(V).3 Teacher College Credit Var.	611.37	589.20	.28	.03	---		.59
G2(M). Mean Classroom Heterogeneity (33.2% of the variance was extracted)							
G2(M).1 Classroom Size	19.03	2.83	---	.07	-.04	.16	.45
G2(M).2 Classroom MPA Heterogeneity	.39	.14	.07	---	.02	.17	.54
G2(M).3 Classroom Homework Heterogeneity	2.12	.40	-.04	.02	---	.20	.46
G2(m).4 Classroom Expect. Heterogeneity	4.87	.85	.16	.17	.20	---	.79
G2(V). Variation of Classroom Heter. (26.8% of the variance was extracted)							
G2(V).1 Classroom Size Var.	62.85	25.85	---	-.01	.04	.02	.55
G2(V).2 Classroom MPA Heter. Variation	.06	.08	-.01	---	-.05	.03	-.47
G2(V).3 Classroom Homework Heter. Var.	5.75	1.88	.04	-.05	---	.02	.71
G2(V).4 Classroom Expect. Heter. Var.	7.50	2.98	.02	.03	.02	---	.21

TABLE 6-A-8

CLUSTER ANALYSIS RESULTS FOR
SCHOOL MEDIATION--SATISFACTION

Cluster		Mean	S. D.	Correlations			Principal Component
H1(M). Mean Attitude Toward Study (51.3% of the variance was extracted)							
H1(M).1	Student Attitude*	-----	----	---	---		---
H1(M).2	Homework	2.96	.36	---	.03		.72
H1(M).3	Parent Attitude	2.00	.14	.03	---		.72
H1(V). Variation of Attitude Toward Study(38.6% of the variance was extracted)							
H1(V).1	Stud. Attitude Var.	.64	.14	---	-.09	.05	.56
H1(V).2	Homework Variation	2.57	.46	-.09	---	-.10	-.69
H1(V).3	Parent Att. Var.	1.03	.12	.05	-.10	---	.61
H2(M). Mean Social Interaction (41.7% of the variance was extracted)							
H2(M).1	Unpleasant Exper.	2.16	.14	---	-.07	.03	-.12
H2(M).2	Teacher Attitude	3.29	.13	-.07	---	.25	.80
H2(M).3	Activities	2.78	.50	.03	.25	---	.77
H2(V). Variation of Social Interaction (47.6% of the variance was extracted)							
H2(V).1	Unpleasant Experience Var.	.68	.16	---	.30	.17	.75
H2(V).2	Teacher Attitude Var.	.68	.19	.30	---	.16	.74
H2(V).3	Activities Var.	2.83	.66	.17	.16	---	.57

*Omitted

Ability. The analytic results for these clusters are presented in Table 6-A-9. All clusters have high and homogeneous interindicator correlations; the correlations are about .70 for mean-level clusters and about .40 for variation clusters, and the principal component loadings are about .90 for mean-level clusters and about .80 for variation clusters. Although this seems high, it is necessary to raise a question in the analyses. At what stage should the test scores have been factored and analyzed. The approach used of clustering school means and variances was dictated by the need to maintain analytic uniformity with the other analyses. However, it is possible that some of the factorial complexity present within the ITED battery could have been better represented had a differently staged analysis been selected. Presumably the factors present in the battery are individualistics; they could better have been detected by dealing with the student by test scale matrix. The factors derived could have been averaged over schools to obtain school scores. The analysis as performed probably blurred the actual complexity of the battery. This will be observed later in the image analysis of the student output cluster.

Aspiration. The analytic results for these clusters are presented in Table 6-A-10. The mean aspiration cluster yields a component which is correlated only with educational aspiration and discrepancy. Neither of those two indicators correlates with girls aspirations. Yet the principal component correlates with the essential concept. The variations in aspiration indicators have lower but more homogeneous correlations. The principal component correlates highly with all three indicators, and thus must be a reasonable representation of this variation.

Emotional Outlooks. The analytic results for this cluster are presented in Table 6-A-11. Both clusters have extremely low interindicator correlations, and, consequently, the principal component analyses failed to provide reasonable composites. These clusters must therefore be omitted from consideration.

B. Factor Structures

The principal component analysis of the clusters was described in the previous section. From each cluster of indicators, a principal component scale was defined. Computationally, the next step was to compute for each school a principal component score for each cluster. This may be considered as a reduction of variables--as a compositing operation. The resulting scores were used as input for image analysis. Separate image analyses were performed for each of the three superclusters: Community Input, School Mediation, and Student Output.

In this section, the results of those analyses are reported. For each analysis, the correlation matrix of the clusters is presented. A table is then presented which contains the major outputs of the image analysis. The anti-image variance of a cluster is the proportions of the

TABLE 6-A-9

CLUSTER ANALYSIS RESULTS FOR
STUDENT OUTPUT--ABILITY

Cluster	Mean	S. D.	Correlations				Principal Component
I1(M). Mean Verbal Ability (80.8% of the variance was extracted)							
I1(M).1 Expression	19.05	1.23	---	.72	.67		.89
I1(M).2 Vocabulary	19.24	1.29	.72	---	.74		.91
I1(M).3 Literature	18.94	1.47	.67	.74	---		.90
I1(V). Variation of Verbal Ability (60.3% of the variance was extracted)							
I1(V).1 Expression Var.	24.51	7.49	---	.39	.37		.75
I1(V).2 Vocabulary Var.	26.73	8.08	.39	---	.45		.80
I1(V).3 Literature Var.	33.21	9.38	.37	.45	---		.79
I2(M). Mean Science and Quantification Ability (78.5% of the variance was extracted)							
I2(M).1 Science Background	20.36	1.42	---	.66	.70		.89
I2(M).2 Quantification Bkgd.	19.39	1.70	.66	---	.68		.88
I2(M).3 Science Interp.	20.00	1.73	.70	.68	---	.90	.90
I2(V). Variation of Science and Quantification Ability (64.0% of variance was extracted)							
I2(V).1 Science Bkgd. Var.	26.46	9.05	---	.46	.49		.82
I2(V).2 Quant. Bkgd. Var.	41.64	10.83	.46	---	.42		.78
I2(V).3 Science Interp.Var.	44.53	12.32	.49	.42	---		.80
I3(M). Mean Social Science Ability (89.2% of variance was ext.,							
I3(M).1 Social Concepts	19.81	1.51	---	.76	.87		.93
I3(M).2 Social Science Int.	19.38	1.73	.76	---	.88		.93
I3(M).3 Inform. Source Util.	21.03	1.59	.87	.88	---		.97
I3(V). Variation of Social Science Ability(76.0% of the variance was extracted)							
I3(V).1 Social Concepts Var.	32.72	9.70	---	.50	.77		.87
I3(V).2 Soc. Sci. Interp.Var.	40.88	10.08	.50	---	.65		.81
I3(V).3 Inf. Source Util.Var.	39.21	10.69	.77	.65	---		.93

TABLE 6-A-10

CLUSTER ANALYSIS RESULTS FOR
STUDENT OUTPUT--ASPIRATION

Cluster	Mean	S. D.	Correlations			Principal Component
J1(M). Mean Aspiration (53.7% of the variance was extracted)						
J1(M).1 Educational Aspiration	6.44	.67	---	.60	.06	.89
J1(M).2 Educational Discrepancy	3.83	.72	.60	---	.06	.89
J1(M).3 Girls' Aspirations	3.40	.30	.06	.06	---	.17
J1(V). Variation in Aspiration (54.2% of the variance was extracted)						
J1(V).1 Educational Aspiration Variation	6.97	1.50	---	.48	.23	.82
J1(V).2 Educational Discrepancy Variation	9.00	.254	.48	---	.20	.80
J1(V).3 Girls' Aspirations Variation	.81	.44	.23	.20	---	.56

TABLE 6-A-11

CLUSTER ANALYSIS RESULTS FOR
STUDENT OUTPUT--EMOTIONAL OUTLOOKS

Cluster	Mean	S. D.	Correlations				Principal Component
K1(M). Mean Adjustment (28.1% of the variance was extracted)							
K1(M).1 Markpoint Variance	2.34	1.55	---	.02	.07	-.05	.51
K1(M).2 Aspirational Discrepancy	.81	.41	.02	---	.02	-.10	.61
K1(M).3 Uncertainty	9.24	6.02	.07	.02	---	.02	.30
K1(M).4 Capacity Learning	.73	.10	-.05	-.10	.02	---	-.63
K1(V). Variation of Adjustment (60.0% of the variance was extracted)							
K1(V).1 Markpoint Variance Variation	18.42	28.55	---	-.04			-.72
K1(V).2 Aspirational Discrepancy Variation	3.72	1.76	-.04	---			.72

cluster score's variation which is not predictable from the other scores. This proportion is related to the relative uniqueness of the score. The Harris roots of G (the image covariance matrix) are also given. The final roots when large, indicate that a poor sample of the content domain has been achieved. The rotated image factor matrix is given. For each factor, a factor sum of squares is given--this is simply the sum of squares of the loadings on the factor, and it indexes the relative importance of the factor. In the accompanying texts, some interpretations and evaluations of the analyses are given and an attempt is made to title the factors. For notation, the codes CI-1, CI-2, ..., CI-8 will be used to identify the community input rotated factors (the order is as given in Table 6-B-2). The codes SM-1, SM-2, ..., SM-7 will be used to identify school mediation rotated factors (the order is as given in Table 6-B-4). The codes SO-1, SO-2, , SO-5 will be used to identify the student output rotated factors (the order is as given in Table 6-B-6).

Community Input

The correlation matrix for the 16 community input cluster scores is given as Table 6-B-1, and the image analysis results are given as Table 6-B-2. Most of the scores have relatively low anti-image variances; i.e., most of the scores share a common variation except for B3 (socioeconomic variation), D4 (farming), EI (archive usage), and E2 (educational involvement). As noted in the previous section, E1 and E2 were unusual in their principal components analysis; E1 consisted of just one probably inaccurate indicator and E2 consisted of two essentially uncorrelated indicators. The last Harris roots are relatively small compared to the first roots. This indicates that the anti-image correlations are small and that a reasonable approximation to a universe of content has been achieved. The rotated image factor matrix contains four important factors, CI-1 to CI-4. The factor structure is simple. Ten of the variables have one large loading on the first four factors; two variables have two large loadings; two factors have one moderate loading, and one variable (E1) has no substantial loadings.

The first factor, CI-1, has high loadings on B2 (socioeconomic level), C1 (social activity/involvement), C2 (social mobility), D2 (industry), D3 (trade), and negatively on B1 (urban). This factor can be identified as urbanization. Since the data are from Iowa, the characteristics of urbanization are related to the particular urbanization patterns of Iowa. For example, there are few big cities, and consequently fewer overwhelming urban problems such as poverty. The people in areas with high scores on this factor tend to have high socioeconomic status, high social mobility and activity levels, and to be involved in industry or trade.

The second factor, CI-2, may be called vitality. The loadings are high with A1 (growth), A2 (youth), A3 (density) and D1 (wealth). The areas with high scores on this factor are achieving population and economic growth.

The third factor, CI-3, is difficult to interpret. It correlates with B3 (socioeconomic variation), D3 (trade), E3 (educational attainment), and negatively with B1 (rural/urban). Also there are five smaller

TABLE 6-B-1

INTERCORRELATIONS* OF COMMUNITY INPUT CLUSTER SCORES

		A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	D4	E1	E2	E3
Growth	A1	100	65	31	-36	58	27	20	17	-45	63	44	25	28	2	10	25
Youth	A2	65	100	62	-15	23	13	-5	2	-30	51	7	-9	0	7	26	4
Density	A3	31	62	100	25	-17	-23	-40	-6	2	20	-33	-50	-14	-14	21	-33
Rural/Urban	B1	-36	-15	25	100	-54	-46	-58	-39	25	-31	-64	-66	-30	-13	-1	-46
Socioeconomic Level	B2	58	23	-17	-54	100	32	47	45	-59	46	75	52	57	7	6	55
Socioeconomic Variation	B3	27	13	-23	-46	32	100	39	3	-35	40	41	49	10	23	1	50
Social Activity/Invmt.	C1	20	-5	-40	-58	47	39	100	42	-12	10	68	63	18	10	-6	49
Social Mobility	C2	17	2	-6	-39	45	3	42	100	12	-6	50	27	26	-3	-9	16
Poverty Culture	C3	-45	-30	2	25	-59	-35	-12	12	100	-63	-32	-32	-34	-14	-8	-41
Wealth	D1	63	51	20	-31	46	40	10	-6	-63	100	27	30	21	18	16	32
Industry	D2	44	7	-33	-64	75	41	68	50	-32	27	100	66	43	9	-11	40
Trade	D3	25	-9	-50	-66	52	49	63	27	-32	30	66	100	30	17	-1	53
Farming	D4	28	0	-14	-30	57	10	18	26	-34	21	43	30	100	4	4	31
Archive Usage	E1	2	7	-14	-13	7	23	10	-3	-14	18	9	17	4	100	10	23
Educational Involvement	E2	10	26	21	-1	6	1	-6	-9	-8	16	-11	-1	4	10	100	35
Educational Attainment	E3	25	4	-33	-46	55	50	49	16	-41	32	40	53	31	23	35	100

*The entries have been multiplied by 100 and rounded.

TABLE 6-B-2

IMAGE ANALYSIS RESULTS*FOR COMMUNITY INPUT

		Rotated Image Factor Matrix With Factor Sums of Squares								Anti-image	Harris
		CI-1	CI-2	CI-3	CI-4	CI-5	CI-6	CI-7	CI-8	Variances	Roots of G
		272	197	187	156	42	3	3	0		
Growth	A1	29	66	15	33	2	3	2	1	.32	16.358
Youth	A2	2	78	2	6	15	0	-1	0	.32	6.169
Density	A3	-25	64	-34	-14	14	-4	-4	-2	.33	2.045
Rural/Urban	B1	-58	-10	-41	-15	2	0	-8	1	.43	1.152
Socioeconomic Level	B2	57	22	22	61	4	3	-2	1	.17	.690
Socioeconomic Variation	B3	22	11	59	12	-0	7	-3	-0	.53	.037
Social Activity/Involvement	C1	65	-10	38	6	-3	7	2	2	.38	.011
Social Mobility	C2	65	1	-13	6	-0	-4	-5	-2	.49	.002
Poverty Culture	C3	-1	-30	-36	-60	-3	-5	-1	-1	.36	.001
Wealth	D1	2	55	38	38	4	2	5	-0	.38	.009
Industry	D2	73	7	28	34	-15	3	2	2	.22	.022
Trade	D3	52	-11	54	24	-7	2	10	1	.33	.043
Farming	D4	32	1	3	49	5	-4	1	-1	.62	.151
Archive Usage	E1	-0	0	30	4	8	-3	-2	0	.87	.318
Educational Involvement	E2	-7	17	10	3	48	-0	-0	0	.65	.589
Educational Attainment	E3	32	-5	51	33	33	7	-1	0	.35	.669

*The values are multiplied by 100 and rounded.

TABLE 6-B-3

INTERCORRELATIONS* OF SCHOOL MEDIATION CLUSTER SCORES

		F1	F2	F3	F4	F5	G1(M)	G1(V)	G2(M)	G2(V)	H1(M)	H1(V)	H2(M)	H2(V)
Mathematics	F1	100	16	-21	7	-9	-5	-8	1	-7	-7	12	-6	2
Science	F2	16	100	6	-29	-9	-0	6	1	-10	-4	12	-5	2
English	F3	-21	6	100	5	13	14	23	-11	-8	-12	13	0	5
History	F4	7	-29	5	100	-10	-21	-13	-8	1	-1	-10	3	-2
Foreign Language	F5	-9	-9	13	-10	100	12	3	4	3	-15	12	-7	4
Mean Teacher Exper.	G1(M)	-5	-0	14	-21	12	100	48	9	-9	-24	28	-10	13
Variation of Teacher Experience	G1(V)	-8	6	23	-13	3	48	100	14	-7	-5	24	-8	18
Mean Classroom Heterogeneity	G2(M)	1	1	-11	-8	4	9	14	100	20	1	-6	-30	21
Variation of Classroom Heterogeneity	G2(V)	-7	-10	-8	1	3	-9	-7	20	100	21	-45	0	14
Mean Attitude Toward Study	H1(M)	-7	-4	-12	-1	-15	-24	-5	1	21	100	-37	31	-26
Variation of Attitude Toward Study	H1(V)	12	12	13	-10	12	28	24	-6	-45	-37	100	-19	25
Mean Social Interaction	H2(M)	-6	-5	0	3	-7	-10	-8	-30	0	31	-19	100	-12
Variation of Social Interaction	H2(V)	2	2	5	-2	4	13	18	21	-4	-26	25	-12	100

*The entries have been multiplied by 100 and rounded.

TABLE 6-B-4

IMAGE ANALYSIS RESULTS*FOR SCHOOL MEDIATION

		Rotated Image Factor Matrix With Factor Sums of Squares							Anti-image Variances	Harris Roots of G
		SM-1 42	SM-2 70	SM-3 53	SM-4 24	SM-5 24	SM-6 2	SM-7 0		
Mathematics	F1	-7	11	-6	-5	-23	-1	-1	.88	1.717
Science	F2	1	10	-2	-28	-8	5	-0	.84	.523
English	F3	14	14	1	-0	25	8	-1	.84	.370
History	F4	-12	1	5	30	-4	5	-1	.82	.230
Foreign Language	F5	3	6	-11	-1	19	-7	-1	.91	.064
Mean Teacher Experience	G1(M)	38	15	-18	-14	18	-5	-1	.68	.022
Variation of Teacher Experience	G1(V)	41	9	-14	-13	17	2	1	.69	.003
Mean Classroom Heterogeneity	G2(M)	10	-20	-32	-6	-3	-1	1	.79	.000
Variation of Classroom Heterogeneity	G2(V)	-6	-44	-2	7	1	-1	-0	.75	.005
Mean Attitude Toward Study	H1(M)	-6	-36	30	2	-9	3	2	.71	.108
Variation of Attitude Toward Study	H1(V)	18	49	-20	-12	5	-1	1	.61	.212
Mean Social Interaction	H2(M)	-3	-8	37	5	0	2	3	.80	.262
Variation of Social Interaction	H2(V)	14	13	-28	-2	3	2	4	.84	.328

*The entries are multiplied by 100 and rounded loadings greater than .25. A reasonable hypothesis would seem to be that areas with high scores on this factor are small urban towns located as trade centers in rural areas. The high variation in socioeconomic standing would then be due to the combination of farmers and tradesmen. The high degree of educational attainment would be due to town-based professional persons serving the surrounding area. The smaller loadings, such as with C1 (social activity/involvement), D1 (wealth), and E1 (archive usage) seem to verify this hypothesis. A reasonable title for the factor might then be trade centers.

The fourth factor, CI-4, may be called wealth where that is intended in the sense of level or standard of living. The high loadings are with B2 (socioeconomic level), D4 (farming), and negatively with C3 (poverty culture).

School Mediation

The correlation matrix for the school mediation cluster scores is given in Table 6-B-3, and the results of the image analysis are given in Table 6-B-4. As noted in the previous section, the measurement of these clusters was imperfect. In the correlation matrix there are only two substantial entries: the correlation between B1 (M) (mean teacher experience) and B1 (v) (variation of teacher experience) is .48 and the correlation between G2 (v) (variation of classroom heterogeneity) and H1 (V) (variation of attitude toward study) is -.45. The anti-image variances are all high, which indicates that most of the variance of each variable is unique, i.e., unrelated to the other variables. The Harris roots are very low, indicating that there was not much commonality. None of the rotated factors have large variances: i.e., none are important. However, in order to satisfy the needs of analysis in the next section, the first five factors were retained.

The first factor, SM-1, has loadings on G1 (M) (mean teacher experience) and G1 (V) (variation of teacher experience). It will be called teacher experience. The second factor, SM-2 has a positive loading on HI (V) (variation of attitude toward study) and negative loadings on G2 (V) (variation of classroom heterogeneity) and H1 (M) (mean attitude toward study). It will be called conflict, which is suggested by the attitude clusters. The third factor, SM-3, has positive loadings with H1 (M) (mean attitude toward study) and H2 (M) (mean social interaction) and negative loadings with G2 (M) (mean classroom heterogeneity) and H2 (M) (variation of social interaction). It will be called comfort, which is intended to imply a pleasant social environment. The fourth and fifth factors are related to the stimulation substance clusters, which were totally obscure, and will not be titled.

Student Output

The correlation matrix for the student output cluster scores is presented as Table 6-B-5 and the results of the image analysis are presented as Table 6-B-6. The intercorrelations for the clusters that

TABLE 6-B-5

INTERCORRELATIONS* OF STUDENT OUTPUT CLUSTER SCORES

		I1(M)	I1(V)	I2(M)	I2(V)	I3(M)	I3(V)	J1(M)	J1(V)	K1(M)	K1(V)
Mean Verbal Ability	I1(M)	100	-25	81	-18	88	-9	36	1	1	-7
Variation of Verbal Ability	I1(V)	-25	100	-28	62	-22	74	-13	1	-22	-0
Mean Science and Quantification Ability	I2(M)	81	-28	100	-30	88	-19	33	1	-0	-3
Variation of Science and Quantification Ability	I2(V)	-18	62	-30	100	-21	80	-6	12	-17	-5
Mean Social Science Ability	I3(M)	88	-22	88	-21	100	-15	34	-0	-4	-7
Variation of Social Science Ability	I3(V)	-9	74	-19	80	-15	100	-10	12	-21	-2
Mean Aspiration	J1(M)	36	-13	33	-6	34	-10	100	-32	-27	-28
Variation of Aspiration	J1(V)	1	1	1	12	-0	12	-32	100	16	9
Mean Adjustment	K1(M)	1	-22	-0	-17	-4	-21	-27	16	100	-13
Variation of Adjustment	K1(V)	-7	-0	-3	-5	-7	-2	-28	9	-15	100

*The entries are multiplied by 100 and rounded.

TABLE 6-B-6

IMAGE ANALYSIS RESULTS*FOR STUDENT OUTPUT

		Rotated Image Factor Matrix With Factor Sums of Squares						
		SO-1 251	SO-2 196	SO-3 35	SO-4 19	SO-5 1	Anti-image Variances	Harris Roots of G
Mean Verbal Ability	I1(M)	88	-8	1	-10	-3	.19	15.757
Variation of Verbal Ability	I1(V)	-17	73	3	4	6	.39	5.765
Mean Science and Quantification Ability	I2(M)	86	-18	4	-4	2	.20	.559
Variation of Science and Quantification Abil.	I2(V)	-14	78	-4	-6	-5	.33	.155
Mean Social Science Ability	I3(M)	91	-11	4	-6	1	.13	.023
Variation of Social Science Ability	I3(V)	-5	84	-4	-1	-1	.23	.006
Mean Aspiration	J1(M)	33	-6	37	-25	-5	.61	.055
Variation of Aspiration	J1(V)	3	9	-33	-12	-3	.84	.083
Mean Adjustment	K1(M)	-4	-24	-32	-5	3	.77	.250
Variation of Adjustment	K1(V)	-4	-2	-6	30	-0	.84	.883

* Entries have been multiplied by 100 and rounded.

were derived from the ITED test battery are generally high and the anti-image variances are correspondingly low. The other clusters have low correlations and high anti-image variances. The last Harris roots are relatively high, which indicates that part of the content domain was not sampled well. Presumably, this part has to do with emotional outlooks of departing seniors. Three rotated factors were retained. The third has a small variance but represents what can be shown of the aspiration clusters.

The first factor, SO-1, loads highly on the three mean-level test battery clusters, I1 (M), I2 (M), and I3 (M) and also has a small loading with J1 (M) (mean aspiration). This factor will be called achievement, intended in its academic sense. It should be noted again that because of the clustering procedure, any factorial structure within the test battery has been smoothed out. However, this factor is doubtless a stable measure of ability or achievement level.

The second factor, SO-2, loads highly on the three variation test battery clusters, I1 (V), I2 (V) and I3 (V). Again, any factorial complexity of the test battery segments has been smoothed out. Schools with high values on this factor have a high variance of scores on the tests: probably many students do rather poorly and many do quite well. This factor will be called academic heterogeneity.

The third factor, SO-3, has, as noted, a very low variance. Yet it is retained because it has loadings with the aspiration clusters. In particular, it loads positively on J1 (M) (mean aspiration) and negatively on J1 (V). (There is also a loading on K1 (M), but that cluster is uninterpretable.) This factor will be called aspiration. Note that schools with high scores on this factor tend to have high aspiration levels and low variance in aspiration.

C. Regression Structures

Analyses of the relationships among the image factors described in the last section and of the relationships between those factors and high school size are presented below. The analyses do not correspond to the full complexity of the theoretical substantive framework described in Section 5-A. That framework consisted of a series of interrelated and directed relationships. Community Input clusters were supposed to represent processes which directly influence the characteristics of students entering high school and which exert a continuing influence over students while they are in school. The Student Output factors were supposed to represent the characteristics--academic, social, and emotional--of the graduating class. The School Mediation factors were supposed to represent the processes occurring in a school which determine the transforming properties of the school and which, along with community processes, produce graduates. Within this context, the single School Manipulation variable--namely, size--was considered to influence the School Mediation processes and in turn to relate to the other clusters.

The School Mediation cluster measurements are critical to such analysis. These measurements, it was noted, were imprecise. Consequently, a simple set of analyses was designed and made. In particular, regression analyses were performed in order to indicate what sorts of Student Output characteristics are independently related to the Community Input and School Mediation factors and to size. In addition, contour maps were prepared to give a better idea of the precise arrangement of some of the relationships. The School Mediation factors were included in both these analyses in spite of the fact that their meanings are still unclear.

The shortcomings in the data and in the earlier analysis meant that this final stage in data analysis was not developed as it might have been. In particular, no substantive conclusions will be offered.

Correlation Analysis

The inter-correlations of the factors and size are presented in Table 6-C-1. Within each supercluster--Community Input, School Mediation, and Student Output--the factors are completely uncorrelated. The correlation matrix is blocked according to size and the three sets of image factors. The largest correlation is between size and urbanization: big schools are usually found in urban places. All the other correlations are quite low--only two are above .30. It should be noted that size correlates moderately with almost all the variables.

Regression Analysis

Within each supercluster the factor scores are automatically uncorrelated. In multiple regression with uncorrelated predictor variables, the amount of variance accounted for by the set of predictors is exactly the sum of the amounts of variance accounted for by the individual predictors. (This is not true for correlated predictors.) Thus, in the present regression analysis, both groups of independent factors could be manipulated as though they were single variables.

Each of the three Student Output factors was considered separately. For each factor, seven regressions were run, corresponding to the seven combinations of the three groups of predictor factors, including size. The results are presented in Tables 6-C-2 to 6-C-4. For each regression equation, there is a column giving the standard regression coefficients. The degree of determination achieved by each equation is summarized in Table 6-C-5. For each equation, the coefficient of determination (squared-multiple-correlation) is presented. All these coefficients are small; no more than 13% of the variance in Student Output is ever achieved. It should be kept in mind, therefore, that in the following discussions of differences in the effect of groups of independent variables, only a small part of the total variation of Student Output is in question.

TABLE 6-C-1

INTERCORRELATIONS OF THE FACTORS PLUS SIZE

		Size	CI-1	CI-2a	CI-3	CI-4	SM-1	SM-2	SM-3	SM-4	SM-5	SO-1	SO-2	SO-3
	Size	1.00	.58	.15	.32	.05	.26	.16	-.47	-.24	.26	.19	.21	-.21
(Urbanization)	CI-1	.58	1.00	0	0	0	.29	.25	-.40	-.15	.16	.03	.05	-.16
(Vitality)	CI-2	.15	0	1.00	0	0	-.13	-.08	-.02	.07	.11	-.03	.08	-.09
(Trade Center)	CI-3	.32	0	0	1.00	0	.12	.20	-.10	-.03	.07	.28	.00	.02
(Wealth)	CI-4	.05	0	0	0	1.00	-.11	.32	-.01	.10	.00	.15	.04	.05
(Teacher Experience)	SM-1	.26	.29	-.13	.12	-.11	1.00	0	0	0	0	.11	.10	-.02
(Conflict)	SM-2	.16	.25	-.08	.20	.32	0	1.00	0	0	0	.17	.08	.03
(Comfort)	SM-3	-.47	-.40	-.02	-.10	-.01	0	0	1.00	0	0	-.07	-.13	.21
(------)	SM-4	-.24	-.15	.07	-.03	.10	0	0	0	1.00	0	.02	.00	.06
(------)	SM-5	.26	.16	.11	.07	.00	0	0	0	0	1.00	.02	.02	.06
(Achievement)	SO-1	.19	.03	-.03	.28	.15	.11	.17	-.07	.02	.02	1.00	0	0
(Academic Heterogen.)	SO-2	.21	.05	.08	.00	.04	.10	.08	-.13	.00	.02	0	1.00	0
(Aspiration)	SO-3	-.21	-.16	-.09	.02	.05	-.02	.03	.21	.06	.06	0	0	1.00

Note: Entries which are automatically zero are printed without a decimal point.

TABLE 6-C-2

REGRESSION WEIGHTS USING
DEPENDENT VARIABLE SO-1
(ACHIEVEMENT)

		Size	CI	SM	Size + CI	Size + SM	CI + SM	Size + CI + SM
	Size	.19			.15	.19		.16
(Urbanization)	CI-1		.03		-.06		-.05	-.11
(Vitality)	CI-2		-.03		-.05		-.01	-.04
(Trade Center)	CI-3		.28		.23		.25	.20
(Wealth)	CI-4		.16		.15		.14	.12
(Teacher Experience)	SM-1			.11		.06	.10	.08
(Conflict)	SM-2			.17		.14	.09	.09
(Comfort)	SM-3			-.07		.02	-.06	-.01
-----	SM-4			.02		.07	.01	.04
-----	SM-5			.02		-.02	.02	-.01

TABLE 6-C-3

REGRESSION WEIGHTS USING
DEPENDENT VARIABLE SO-2
(ACADEMIC HETEROGENEITY)

		Size	CI	SM	Size + CI	Size + SM	CI + SM	Size + CI + SM
	Size	.21			.32	.19		.30
(Urbanization)	CI-1		.05		-.14		-.11	-.22
(Vitality)	CI-2		.08		.03		.10	.06
(Trade Center)	CI-3		.00		-.10		-.06	-.14
(Wealth)	CI-4		.04		.02		.01	-.01
(Teacher Experience)	SM-1			.10		.05	.15	.11
(Conflict)	Sm-2			.08		.04	.12	.12
(Comfort)	SM-3			-.13		-.04	-.18	-.09
-----	SM-4			.00		.05	-.02	.04
-----	SM-5			.23		-.03	.03	-.02

TABLE 6-C-4

REGRESSION WEIGHTS USING
DEPENDENT VARIABLE SO-3
(ASPIRATION)

		Size	CI	SM	Size + CI	Size + SM	CI + SM	Size + CI + SM
	Size	-.21			-.22	-.21		-.21
(Urbanization)	CI-1		-.16		-.03		-.10	-.03
(Vitality)	CI-2		-.09		-.06		-.10	-.06
(Trade Center)	CI-3		.02		.09		.02	.08
(Wealth)	CI-4		.05		.07		.04	.06
(Teacher Experience)	SM-1			-.02		.04	.00	.04
(Conflict)	SM-2			.03		.06	.03	.03
(Comfort)	SM-3			.21		.11	.17	.10
-----	SM-4			.06		.01	.04	.00
-----	SM-5			.06		.11	.08	.12

TABLE 6-C-5

SUMMARY OF THE COEFFICIENTS OF
DETERMINATION FOR THE REGRESSION ANALYSES

Factors	Dependent Variables		
	Achievement SO-1	Academic Heterogeneity SO-2	Aspiration SO-3
Size	.037	.045	.045
Community Input	.105	.010	.035
School Mediation	.045	.034	.050
Size and Community Input	.118	.066	.061
Size and School Mediation	.065	.055	.075
Community Input and School Mediation	.120	.053	.069
Community Input, Size, and School Mediation	.130	.090	.088

Examining the first column of Table 6-C-5, it can be seen that the Community Input factors account for most of the variance that can be accounted for in SO-1 (achievement). Although size and School Mediation do have some effect, most of their effect overlaps with that of Community Input. Community Input accounts for 10.5% of the variance of achievement. The addition of size and School Mediation to the equation increases the percentage to 13.0%. The net increase is 2.5% which is less than half of the total of 6.5% that size and School Mediation account for. Looking at Table 6-B-2, it can be seen that factor CI-3 (trade center) accounts for much of Community Input's effect. Also, CI-4 (wealth) is related. It can also be seen that size does have an effect. The last column of Table 6-B-2 shows a regression coefficient of .16 for size in the equation including all the variables.

The second column in Table 6-C-5 gives the coefficients of determination in the prediction of SO-2 (academic heterogeneity). Maximum coefficient (9%) is about a third smaller than maximum coefficient (13%) for the previous factor. Also, Community Input is a suppressor variable group for both size and School Mediation. Size and Community Input account together for more than the sum of what they account for separately. Exactly which group has the greatest impact on academic heterogeneity is not clear. In Table 6-B-3 it can be seen that SM-5 has a high positive weight when only School Mediation is included, but it has a very small weight when other factors are included. Other reversals seem to take place and the regression structure is undecipherable.

In the third column of Table 6-C-5, it is seen that 8.8% of the variance is accounted for by all the factors in predicting SO-3 (aspiration). Each factor accounts for about the same amount of variance, and the accumulation is approximately linear. Each factor, including size, seems to account for some unique variation. Size and the two untitled School Mediation clusters seem to have the greatest effect.

Contour-Plotting Analysis

The correlations found in the regression analysis were disappointingly low and some contour-plotting checks were made to determine whether non-linear relations or mixture distributions were depressing the correlations. The contour-plotting theory and operations were explained in Section 5-D.

In Figure 6-A-1 is presented a contour plot of the joint frequency distribution of CI-1 (urbanization) and CI-2 (vitality). By their construction, these factors have a correlation of zero. But it is clear from the plot that they do not have a bivariate normal distribution. This might be a mixture distribution, with a positively correlated part (seen in the inner lines) and a negatively correlated part (seen in the outer lines). If that is true, the use of the two variables as two independent dimensions of community character is misleading.

A better example of a mixture distribution with confusing correlation is given in Figure 6-C-2. This is the contour plot of size (vertically with high values at the bottom) and SM-3 (comfort). There are

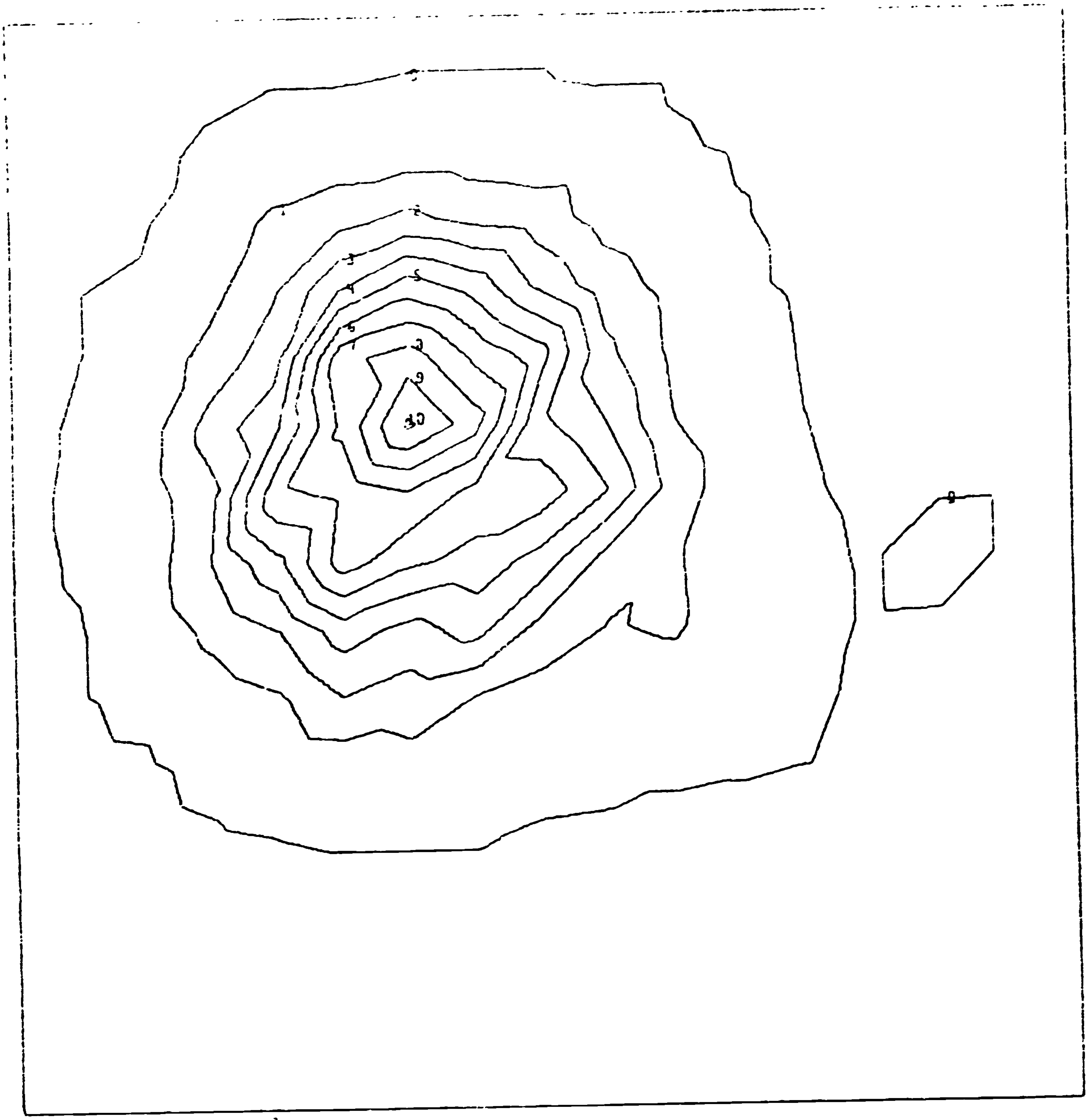


FIGURE 6-C-1

CONTOUR PLOT OF THE JOINT FREQUENCY
DISTRIBUTION OF FACTORS
CI-1 AND CI-2

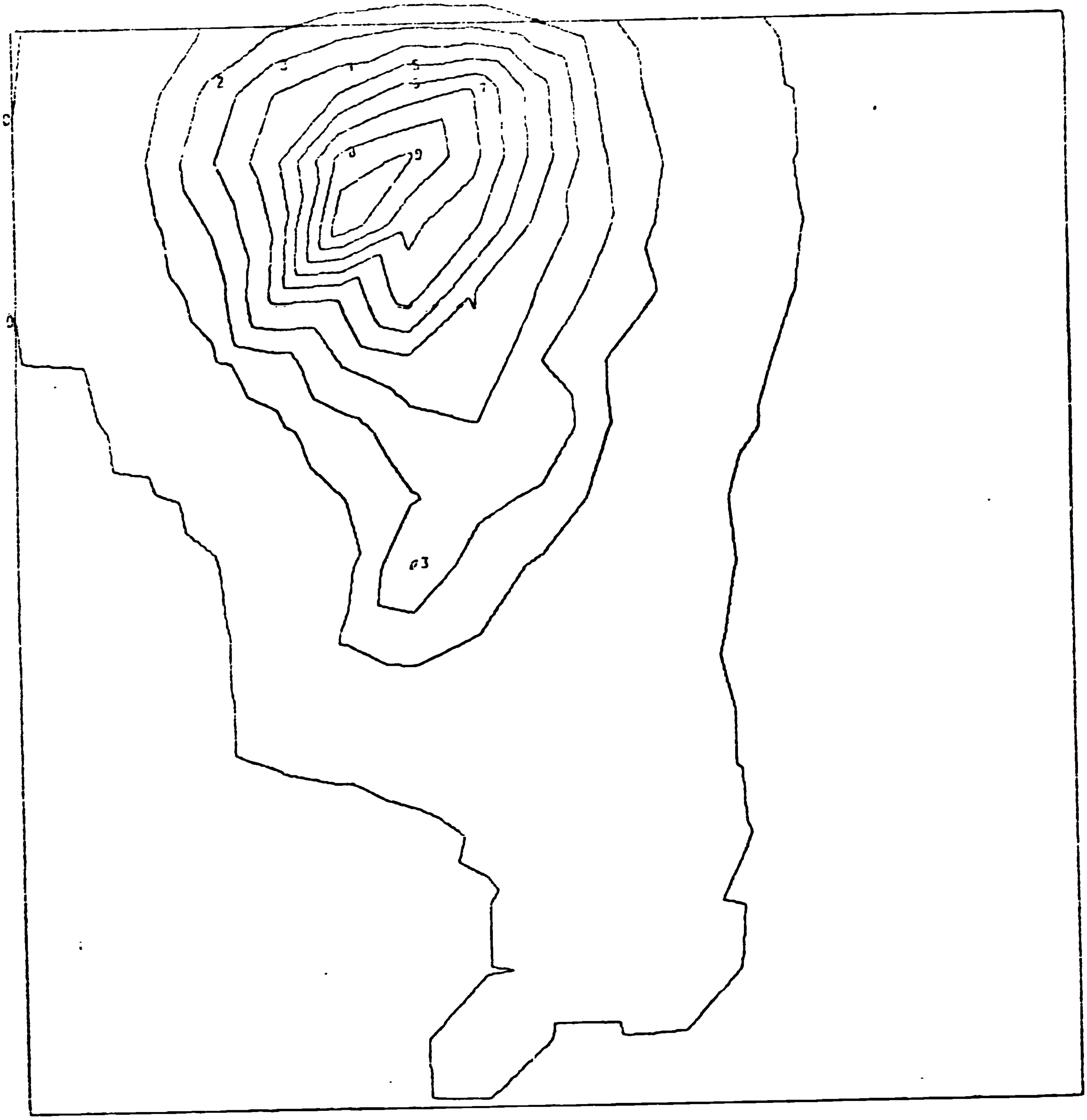


FIGURE 6-C-2

CONTOUR PLOT OF THE JOINT FREQUENCY
DISTRIBUTION OF SIZE AND
FACTOR SM-3

clearly two trends: for small schools the trend is lower-left to upper-right and for large schools, the trend is lower-right to upper-left. These trends are contradictory and the correlation is therefore depressed.

These two examples suffice to illustrate the utility of the countour-plotting approach developed during the project.

D. Concluding Statement

The purpose of this research was to relate the size of secondary schools to certain costs which are not being considered when new educational facilities are planned and built. These costs were measured in terms of educational productivity, which in turn was measured along a wide spectrum of outcomes ranging from academic progress to social behavior.

Three long-range objectives were initially proposed: (1) to determine the interrelationships of community, school, and educational output factors as influenced by size of secondary school; (2) to apply these interrelationships to a representative group of educational situations; and (3) to disseminate findings to school officials for use in making decisions about size of secondary schools.

This report covers phase one of the project. The specific objective of this phase has been to develop a suitable methodology for subsequent application to a national sample. The population selected for study in phase one consisted of the public secondary schools of Iowa and Florida. The sample was subsequently narrowed to the high schools of Iowa.

A theoretical framework was constructed, into which selected variables were set. Since many of these variables were not directly measurable, indicators were drawn from the information found in data banks. The data were clustered according to the theoretical framework and analyses were then performed.

Five conclusions may be drawn from the work completed in phase one:

1. It may be less costly in time and money to gather data for intended research uses than to attempt to convert or adapt data gathered for other purposes in order to generalize the solution of operational problems.

2. The compiling of data on social characteristics of educational units (attendance or administrative) has generally been neglected by all levels of government, including school districts. The data which are maintained have immediate and practical use in matters of management, but they are insufficient in scope and often in unsuitable form for research aimed at solutions to problems of a general nature. Yet, these social characteristics are becoming more and more important to know and consider in educational planning.

3. The ten factors extracted in this research may be useful in describing principal elements in the process by which community inputs become student outputs via school mediation. However, they probably do not yet provide a reasonably complete description of the social phenomena because of the character and coverage of the data used.

4. Size of secondary school was found to be significantly related to one factor which is itself a function of size and density, namely urbanization.

5. The theoretical framework, as originally constructed, was intended to describe the relationship in its complexity. Yet, as has been indicated, the data were inadequate for proper testing of the theory and modifications had to be made in the clustering of the data. These modifications contributed to the reduction in value of principal component analysis.

In the next phase of this research, a national sample of secondary schools will be drawn. Data appropriate to the indicators set in the theoretical framework will then be gathered by direct means and analyzed by the method developed in this first phase. A general answer can then be given to the consequences of high school size under investigation in this project.

REFERENCES

Articles

1. Aiken, Lewis R., Jr., "Rank in High School Graduating Classes of Various Sizes as a Predictor of College Grades," Journal of Educational Research. 58 (2) (October, 1964), pp. 56-60.

2. Allen, Francis R. and Bentz, W. Kenneth, "Toward the Measurement of Sociocultural Change," Social Forces. 43 (1965), pp. 523-32.

3. Anderson, Theodore R. and Bean, Lee L., "The Shevky-Bell Social Areas: Confirmation of Results and a Reinterpretation," Social Forces. 40 (1961), pp. 119-21.

4. Anderson, Theodore R. and Egeland, Janice A., "Spatial Aspects of Social Area Analysis," American Sociological Review. 26 (1961), pp. 392-99.

5. Anderson, T. R. and Seymour, Warkov, "Organizational Size and Functional Complexity: A Study of Administration in Hospitals," American Sociological Review. 26, (1961), pp. 23-8.

6. Argyris, Chris, "The Individual and Organization: An Empirical Test," Administrative Science Quarterly. 4, (1959), pp. 145-67.

7. Argyris, Chris, "We Must Make Work Worthwhile!" Life Magazine, Vol. 62, No. 18 (May 5, 1967), pp. 56-68.

8. Babchuk, Nicholas, "Work Incentives in a Self-Determined Group," American Sociological Review. 16, (1951), pp. 679-87.

9. Baumgartel, Howard, "Leadership? Motivation and Attitudes in Research Laboratories," Journal of Social Issues, 12, (1956), pp. 24-31.

10. Baumgartel, Howard, "Leadership Styles as a Variable in Research Administration," Administrative Science Quarterly. 2, (1957), pp. 344-60.

11. Bell, Wendell, "Economic, Family, and Ethnic Status: An Empirical Test," American Sociological Review. 20 (1955), pp. 45-52.

12. Bell, Wendell, "Social Areas: A Typology of Urban Neighborhoods," in Sussman, Marvin (ed.) Community Science and Analysis. New York: Thomas Crowell Company, (1959).

13. Bell, Wendell, "The Utility of the Shevky for the Design of Urban Sub-Area Field Studies," Journal of Social Psychology, 47 (1959), pp. 71-83.

14. Bell, Wendell and Force, Maryanne T., "Social Structure and Participation in Different Types of Formal Associations," Social Forces. 84 (1955), pp. 345-50.

15. Beshers, James M., "Statistical Inferences from Small-Area Data," Social Forces, 38 (May, 1960), pp. 341-48.

16. Boocock, Sarane S., "Toward a Sociology of Learning: A Selective Review of Existing Research," Sociology of Education, 39 (1) (Winter, 1966), pp. 1-45.

17. Buder, Leonard, "Study Here Finds School Officials Mired in Inertia " New York Times, Vol. CXVI No. 40,013, (August 13, 1967), pp. 1-41.

18. Buechley, R. W., "Review of Social Area Analysis," Journal of the American Statistical Association. 51 (1956), pp. 145-97.

19. Caplow, Theodore, "Organizational Size," Administrative Science Quarterly. (1957), pp. 484-505.

20. Carlson, Richard, "Succession and Performances Among School Superintendents," Administrative Science Quarterly, 6, (1961), pp. 210-27.

21. Chapin, F. Stuart, "The Optimum Size of Institutions, A Theory of the Large Group," American Journal of Sociology, 92, (1957), pp. 449-66.

22. Clements, Forrest E., "The Use of Cluster Analysis with Anthropological Data," American Anthropologist, 56 (April, 1954), pp. 180-99.

23. Duncan, Otis D., "Research on Metropolitan Population: Evaluation of Data," American Statistical Association Journal. 51 (1956), pp. 591-6.

24. Duncan, Otis D. and Duncan, Beverly, "A Methodological Analysis of Segregation Indexes," American Sociological Review, 20 (April, 1955), pp. 210-17.

25. Duncan, Otis D. and Schnore, Leo F., "Cultural, Behavioral, and Ecological Perspectives in the Study of Social Organization," The American Journal of Sociology, 65 (1959), pp. 132-53.

26. Faber, Charles F., "Measuring School District Quality," The American School Board Journal, (October, 1964), pp. 12-13.

27. Faber, Charles F., "The Size of a School District," Phi Delta Kappan, XLVII (September, 1966), pp. 33-35.

28. Farber, B., and Osoinack, J. C., "An Index of Socio-Economic Rank of Census Tracts in Urban Area," American Sociological Review. 24 (1959), pp. 630-40.

29. Foley, Donald L., "Census Tracts and Urban Research," American Statistical Association Journal. 48 (1953), pp. 733-42.

30. Gibbs, J. R., "The Effects of Group Size and of Threat Reduction upon Creativity in a Problem-Solving Situation," American Psychologist. 6 (1951), p. 324.

31. Gillette, John M., "Community Concepts," Social Forces. 4 (1926), pp. 677-89.

32. Goodman, Leo A., "Some Alternatives to Ecological Correlation," The American Journal of Sociology. LXIV (May, 1957), pp. 610-625.

33. Gouldner, Alvin W., "Cosmopolitan and Locals: Toward Analysis of Latent Social Roles," Administrative Science Quarterly. 2, (1957) pp. 281-306.

34. Greer, Scott, "Urbanism Reconsidered: A Comparative Study of Local Areas in a Metropolis," American Sociological Review. 21 (February, 1956), pp. 19-25.

35. Gross, Edward, "Some Functional Consequences of Primary Controls in a Formal Work Organization," American Sociological Review. 18 (1953), pp. 368-72.

36. Grusky, Oscar, "Corporate Size Bureaucratization and Managerial Successions," American Journal of Sociology. 67 (1962), pp. 261-9.

37. Hamilton, De Forest, and Rowe, Robert N., "Academic Achievement of Students in Reorganized and Non-Reorganized Districts," Phi Delta Kappan. XLIII (June, 1962), pp. 401-4.

38. Hanson, Robert C., "Administrator Responsibility in Large and Small Hospitals Metropolitan Community," Journal of Health and Human Behavior. B Fall, (1964), pp. 199-204.

39. Harmon, Lindsey R., "High School Backgrounds of Science Doctorates," Science. 133 (1961), pp. 679-88.

40. Harris, Chauncey D., and Ullman, Edward L., "The Nature of Cities," in Hatt, Paul, and Reiss, Albert J., Jr. (eds.). Reader in Urban Sociology. Glencoe: The Free Press, (1951), pp. 227-32.

41. Hatt, Paul, "The Concept of Natural Areas," in Thecdorson, George W., (ed.). Studies in Human Ecology. Evanston, Illinois: Row, Peterson and Company, (1961), pp. 104-8.

42. Havie, Mason, "Size, Shape and Function in Industrial Organizations," Human Organization. 14 (1955), pp. 17-22.

43. Hawley, Amos, "Ecology and Human Ecology," Social Forces. 23 (May, 1944), pp. 398-405.

44. Hawley, Amos and Duncan, Otis S., "Social Area Analysis: A Critical Approach," Land Economics. XXXIII (November,1957), pp. 337-44.

45. Herbst, P. G., "Measurement of Behavior Structures by Means of Input-Output Data," Human Relations. 10 (1957), pp. 335-45.

46. Hill, Mozell C., and Whiting, Albert N., "Some Theoretical and Methodological Problems in Community Studies," Social Forces. 29 (1950), pp. 117-23.

47. Hillery, George A., "Definitions of Community: Areas of Agreement," Rural Sociology. 20 (1955), pp. 111-23.

48. Hollingsworth, Paul M., "The Issue of Class Size," Education. 84 (March, 1964), pp. 433-6.

50. Hoyt, Donald P., "Size of High School and College Grades," Personnel and Guidance Journal. 37 (April, 1959), pp. 569-73.

51. Indik, B. P. "The Relationship Between Organization Size and Supervision Ratio," Administrative Science Quarterly. 9 (37 1964), pp. 301-12.

52. Indik, B. P., "Some Effects of Organization Size on Member Attitudes and Behavior," Human Relations. 16 (4) (1963), pp. 369-84.

53. Iowa State University of Science and Technology, "Iowa's Public Schools - A Review," Cooperative Extension Service. Ames, Iowa.

54. Jonassen, Christen T., "Functional Unities in Eighty-Eight Community System," American Sociological Review. 26 (1961), pp. 399-407.

55. Jonassen, Christen T., "Toward an Operational Definition of Community Welfare," Social Forces. 37 (1958), pp. 112-118.

56. Kahl, Joseph and Davis, James A., "A Comparison of Indexes of Socioeconomic Status," American Sociological Review. 20 (June, 1955), pp. 317-25.

57. Kriesberg, "Careers, Organization Size and Succession," American Journal of Sociology. 68 (1962), pp. 355-9.

58. Lennon, Roger T., "The Relation between Intelligence and Achievement Test Results for a Group of Communities " Journal of Educational Psychology. XLI (1950), pp. 301-8.

59. Lindsay, Harmon, "Field of Doctorate Specialization as a Function of Size of High Schools," Science. 130 (1959), p. 1473.

60. Livingston, A. Hugh, "Is There an Optimum Size High School?" Progressive Education. 33 (September, 1956), pp. 156-9.

61. Mabry, John H., "Census Tract Variation in Urban Research," American Sociological Review. 23 (April, 1958), pp. 193-96.

62. Maller, J. B. "Economic and Social Correlatives of School Progress in New York City," Teachers College Record. XXXIV (1933), pp. 655-70.

63. Marriott, R., "Size of Working Group and Output," Occupational Psychology. 23 (1949). pp. 47-56.

64. Mayo, S. S., "What Size High School?" American School Board Journal. 144 (January, 1962), pp. 32-33.

65. McDill, Edward L., and Coleman, James, "Family and Peer Influences in College Plans of High School Students," Sociology of Education. 38 (Winter, 1965), pp. 112-26.

66. Medalia, Nahum Z., "Unit Size and Leadership Perceptions," Sociometry. 17 (1954), pp. 64-7.

67. Menzel, Herbert, "Comment on Robinson's Ecological Correlations and the Behavior of Individuals," in Theodorson, George (ed.). Studies in Human Ecology. Evanston, Illinois: Row, Peterson and Company, (1961), pp. 121-22.

68. Nam, Charles B., "Family Patterns of Educational Attainment," Sociology of Education, 38 (Fall, 1965), pp. 393-403.

69. Peters, William S., "A Method of Deriving Geographic Patterns of Associated Demographic Characteristics Within Urban Areas," Social Forces. 35 (1956), pp. 62-68.

70. Pierce, Truman M., Controllable Community Characteristics Related to the Quality of Education. Research Report No. 1. Prepared by the Metropolitan School Study Council. New York: Bureau of Publications, Teachers College, Columbia University, (1947).

71. Reiss, Albert J. Jr., "An Analysis of Urban Phenomena," in Fisher, Robert M. (ed.). The Metropolis in Modern Life. Garden City: Doubleday and Co., (1955) pp. 41-49.

72. Robinson, W. S., Bell, Wendell, and Brown, Leonard, "Urban Analysis by Social Areas: A Comparison and Evaluation," Unpublished Manuscript.

73. Schloerke, Wallace C., "Does Bigness Insure Quality?" Michigan Education Journal. 42 (May, 1965), pp. 20-21.

74. Schmid, Calvin F., "Generalizations Concerning the Ecology of the American City American Sociological Review, 15 (April, 1950), pp. 264-81.

75. Schnore, Leo F., "Another Comment on Social Area Analysis," Pacific Sociological Review, 5 (1962), pp. 13-15.

76. Schutz, Richard E., "A Factor Analysis of Academic Achievement and Community Characteristics," Educational and Psychological Measurement, 20 (1960), pp. 513-18.

77. Schutz, Richard E., "A Factor Analysis of Educational Development in the United States," Educational and Psychological Measurement. 16 (1956), pp. 325-32.

78. Simmel, George, "The Number of Members as Determining the Sociological Firm of the Group I," American Journal of Sociology, 8 (1902), pp. 1-46.

79. Simmel, George, "The Number of Members as Determining the Sociological Firm of the Group II," American Journal of Sociology, 8 (1902), pp. 158-96.

80. Simion, Herbert, "The Compensation of Executives," Sociometry, 20 (1957), pp. 32-5.

81. Simpson, Richard L. and William H. Gulley "Goals, Environmental Pressures and Organizational Characteristics," American Sociological Review, 27 (1962) pp. 344-50.

82. Smith, Paul M., Jr., "The Large or Small High School?" Journal of Secondary Education, 36 (7) (November, 1961) pp. 389-92.

83. Strauss, George and Leonard R. Sayees, "Organization and the Selection of Local Union Officers," American Journal of Sociology, 58 (1952), pp. 585-91.

84. Talacchi, Sergio, "Organization Size, Individual Attitudes and Behavior: An Empirical Study," Administrative Science Quarterly. 5 (December, 1960) pp. 399-420.

85. Talacchi, Sergio, "Organizational Size, Individual Attitudes and Behavior: An Empirical Study," Administrative Science Quarterly, 5 (1961), pp. 398-420.

86. Terrien, F. W. and Donald L. Mills, "The Effect of Changing Size upon the Internal Structure of Organizations," American Sociological Review. 20 (1955), pp. 11-13.

87. Thomas, Edwin J., "Role Conceptions and Organizational Size," American Sociological Review, 21 (1959) pp. 30-7.

88. Thorndike, Robert L., "Community Variables as Predictors of Intelligence and Academic Achievement," The Journal of Educational Psychology. 42 (1951), pp. 321-38.

89. Thorndike, Robert L., Growing Points in Educational Research. A 1949 Official report of the American Educational Research Association on Community Factors Related to Intelligence and Achievement of School Children. Washington: American Educational Research Association, (1949).

90. Thorndike, Edward L. and Woodyard, Ella, "Differences within and between Communities in the Intelligence of the Children," The Journal of Educational Psychology. XXXIII (December, 1942), pp. 641-56.

91. Tiebout, Charles M., "Hawley and Duncan on Social Area Analysis: A Comment," Land Economics. XXXIV (May, 1958), pp. 182-84.

92. Turner, A. N., "Interaction and Sentiment in the Foreman-Worker Relationship," Human Organization. 14 (1955), pp. 10-16.

93. Van Arsdol, Maurice D. et al., "An Investigation of the Utility of Urban Typology," Pacific Sociological Review. 4 (1961), pp. 26-32.

94. Van Arsdol, M. D., "Further Comments on the Utility of Urban Typology," Pacific Sociological Review. 5 (1957), pp. 9-13.

95. Van Arsdol, Maurice D. et al., "The Generality of Urban Social Area Indexes," American Sociological Review. 23 (1958), pp. 277-84.

96. Van Arsdol, Maurice D., Camilleri, Santo F., and Schmid, Calvin F., "An Application of the Shevky Social Area Indexes to a Model of Urban Society," Social Forces. 37 (1959), pp. 26-32.

97. Warner, W. Keith and James S. Hilander, "The Relationship between Size of Organization and Membership Participation," Rural Sociology. 29 (1964), pp. 30-9.

98. Weiss, Robert S. and Eugene Jacobson, "A Method for the Analysis of the Structure of Complex Organizations," American Sociological Review, 20 (1955), pp. 661-8.

99. Worthy, James C., "Organizational Structure and Employee Morale," American Sociological Review. (1950), pp. 169-79.

100. Young, Nathan, "Community Predictors of School Holding Power," Journal of Secondary Education. 39 (March, 1964), pp. 116-121.

101. Zorbaugh, Harvey W., "The Natural Area of the City," in Theodorson, George W. (ed.). Studies in Human Ecology. Evanston, Illinois: Row, Peterson and Company, (1961), pp. 45-9.

Books

1. American Psychological Association, 1961, American Psychological Association, Washington, D. C., (1961). (APA)

2. Argyris, Chris, Executive Leadership. New York, Harper, (1953).

3. Barker, Roger G., and Gump, Paul V. Big School, Small School: High School Size and Student Behavior. Stanford: Stanford University Press, (1964).

4. Beshers, James M. (ed.). Computer Methods in the Analysis of Large-Scale Social Systems. Cambridge: Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, (1965).

5. Beshers, James M. Urban Social Structure. Glencoe, Illinois: The Free Press on Glencoe, Inc., (1962).

6. Bogue, Donald J. (ed.). Needed Urban and Metropolitan Research. ("Scripps Foundation Studies on Population Distribution," No. 7) Oxford, Ohio: Scripps Foundation, Miami University, (1953).

7. Bogue, Donald J. (ed.). The Structure of the Metropolitan Community: A Study in Dominance and Subdominance. Ann Arbor: University of Michigan, (1950).

8. Boulding, Kenneth E., The Organizational Revolution. Harper and Brothers, New York, (1953).

9. Burgess, Ernest and Bogue, Donald (eds.). Contributions to Urban Sociology. Chicago: University of Chicago Press, (1964).

10. Burling, Temple, Edith Lentz, and Robert N. Wilson, The Give and Take in Hospitals. New York, Putnam, (1956).

11. Cartwright, Darwin and Alvin Zander, Group Dynamics Research and Theory. Row, Peterson and Company, Elmford, New York, (1960).

12. Charters, W. W. and Gage, N. L. (eds.). Readings in the Social Psychology of Education. Boston: Allyn and Bacon,(1963).

13. Cummings and Makintosh, Curriculum Responsibility of State Department of Education. U.S. Office of Education, Miscellaneous Number 30, Washington, D. C., (1960). (CR)

14. Digest of Educational Statistics, 1964, U.S. Government Printing Office, Washington, D. C., (1964). (DES)

15. Duncan, Otis D. and Reiss, Albert J. Social Characteristics of Urban and Rural Communities. 1950. New York: John Wiley and Sons, Inc., (1956).

16. Flanagan, John D., et al. Design for a Study of American Youth. Boston: Houghton Mifflin, (1962).

17. Gaumnitz, Walter H. and Tompkins, Ellsworth, Holding Power and Size of High Schools. FSA Circular No. 322. Washington, D.C.: USGPO, (1950), 25 pp.

18. Gibbs, Jack P. (ed.). Urban Research Methods. Princeton: D. Van Nostrand Company, (1961).

19. Gittell, Marilyn, Participants and Participation: A Study of School Policy in New York City. New York, Center for Urban Education, (1967).

20. Goldberger, Arthur S., Econometric Theory. John Wiley & Sons, Inc., (1964).

21. Government Affairs Foundation, Inc. Metropolitan Surveys: A Digest. Chicago: Public Administration Service, (1958).

22. Gross, Bertram M., The Managing of Organizations, The Free Press of Glencoe, Collier-Macmillan, Limited, London, pp. 357-60, 893-949.

23. Gulick, Luther, "Notes of the Theory of Organization," Papers on the Science of Administration. Institute of Public Administration, Columbia University, New York, (1937).

24. Hadden, Jeffrey K., and Borgatta, Edgar F., Social Characteristics of American Cities. Rand McNally, (1966).

25. Hatt, Paul and Reiss, Albert J. Jr. (eds.). Reader in Urban Sociology. Glencoe, Illinois: The Free Press, (1951).

26. Hauser, Philip M. (ed.). Handbook for Social Research in Urban Areas. Belgium: UNESCO, (1964).

27. Havie, Mason, "Biological Models and Empirical Histories of the Growth of Organizations," Modern Organizational Theory. New York, John Wiley, (1959).

28. Hawley, Amos H. The Changing Shape of Metropolitan America. Glencoe, Illinois: The Free Press, (1956).

29. Hollander, Edwin Paul, Leaders, Groups, and Influence. New York, Oxford University Press, (1964).

30. Jonassen, Christen T. The Measurement of Community Dimensions and Elements. Columbus: Ohio State University Center for Educational Administration, (1958).

31. Kershaw, Joseph A., and McKean, R. N. Systems Analysis and Education. Santa Monica: The Rand Corporation, (1959).

32. Kowitz, Gerald T., and Sayres, William C., Size, Cost and Educational Opportunity in Secondary Schools. New York: The State Education Department Division of Research, New York State University, (1959).

33. Likert, Rensis, "Voluntary Organizations," New Patterns of Management. Chapter 10, New York, McGraw-Hill, (1961).

34. National Education Association. Efficient School Size. Research Memo, 1959-1. Washington: N.E.A., (1959).

35. Ohio Education Association. The Relationship Between Academic Achievements of Students in Colleges and the Size of High Schools from which they were Graduated. Report of the Educational Council, 1959. Columbus, Ohio: The Ohio School Survey Committee, (1959).

36. Park, Robert. Human Communities. Glencoe: The Free Press, (1952).

37. Ranking of the States. National Educational Association, Washington, D.C., (1966). (RS)

38. Reiss, Albert J., Duncan, Otis, Hatt, Paul, and North, Cecil. Occupations and Social Status. New York: The Free Press of Glencoe, Inc., (1961).

39. Reiss, Albert J., Jr. The Metropolis in Modern Life. Garden City: Doubleday and Co., (1955).

40. Reissman, Leonard. The Urban Process. Glencoe, Illinois: The Free Press, (1964).

41. Revans, R. W., "Human Relations, Management and Size," Human Relations and Modern Management. Edward M. Hugh-Jones, editor, Quandrangle Books, Chicago, (1959).

42. Rubenstein, Albert H. and Chadwick, J. Haberstrah, Home Theories of Organization. Dorsey Press, Homewood, Illinois, (1966).

43. Shevky, Eshreff, and Bell, Wendell. Social Area Analysis. Stanford: Stanford University Press, (1955).

44. Shevky, Eshreff, and Williams, Marilyn. The Social Areas of Los Angeles: Analysis and Typology. Berkeley: University of California Press, (1949).

45. Statistics of Education in the U.S., 1958-9 Series, Public Secondary Schools Number 1. Washington, D. C , (1961). (PSS)

46. Statistics of State School Systems: 1955-6, Chapter 2. U.S. Government Printing Office, Washington, D.C., (1959). (SSS)

47. Steward, Julian H. Area Research: Theory and Practice. SSRC Bulletin 63. New York: Social Science Research Council, (1950).

48. Sussman, Marvin B. (ed.). Community Structure and Analysis. New York: Thomas Y. Crowell Company, (1959).

49. Theodorson, George A. (ed.). Studies in Human Ecology. Evanston, Illinois: Row, Peterson and Company, (1961).

50. Trist, E. L., G. W. Higgin, H. Murray and A. B. Pollock, Organizational Choice, London, Taristock, (1963).

51. Tyron, Robert C. Identification of Social Areas by Cluster Analysis. Berkeley: University of California Press, (1955).

52. The U.S. Book of Facts Statistics and Information. 1967, Washington Square Press, Inc., New York, (1966). (USB)

53. U.S. Bureau of the Census, Statistical Abstract of the U.S.. Washington, D.C., (1953, 1961, 1962, 1963). (SA)

54. Broom, Victor H., Some Personality Determinants of the Effects of Participation. Englewood Cliffs, New Jersey, Prentice-Hall, (1960).

55. Weiss, Robert S., Process of Organization. Ann Arbor, Survey Research Center, University of Michigan, (1956).

56. Wilson, Geoffrey and Wilson, Monica. Analysis of Social Change. London: Cambridge University Press, (1949).

57. Wood, Robert C., 1400 Governments: The Political Economy of the New York Metropolitan Region. Cambridge: Harvard University Pres , (1961).

Dissertations

1. Andrews, Lloyd Nelson, Ed.D. "Relationship of High School Size to School-Community Relations." (L.C. Card No. Mic 58-3577) Stanford University, (1958).

2. Barnard, Harry Vollie, Ed.D. "The Relationship of School Size and Accreditation to Certain Factors in Alabama's White Public Accredited Six-Year Secondary Schools." (L.C. Card No. Mic 60-1725) University of Alabama, (1959).

3. Basler, David D., Ph.D. "An Investigation of Certain Factors Influencing the Optimum Size for Elementary School Attendance Units." (L.C. Card No. 60-5639) Iowa State University, (1960).

4. Berg, Arthur David, Ph.D. "The Determination of the Discrimination and Reliability Indices of the Educational Characteristics Criterion with Implications Concerning Educational Cost-Quality Relationship." (L.C. Card No. 63-6141) Michigan State University, (1963).

5. Brown, William Earl, Ed.D. "High School Size: Its Relationship to Selected Educational and Cost Factors." University of Southern California, (1956).

6. Bryant, Milford Howard, Ed.D. "Factors Affecting the Decision of Top Quarter Virginia High School Graduates not to go to College." (L.C. Card No. 24-242) University of Virginia, (1957).

7. Campbell, Joe Watkins, Ed.D. "Factors Related to Scholastic Achievement: Louisiana State University's 1963-1964 Freshman Class." (L.C. Card No. 66-721) Louisiana State University, (1965).

8. Collinsworth, B. Jack, Ed.D. "An Analysis of the Relationship of Size of Arkansas High Schools to Selected Qualifications of High School Teaching Personnel. (L.C. Card No. Mic 61-1337) University of Arkansas, (1961).

9. Conrad, Marion James, Ph.D. "A Technique for Determining the Operating Capacity of Secondary-School Buildings. (L.C. Card No. 23-682) Ohio State University, (1952).

10. Crocker, Jack W., Ed.D. "The Relationship of Size and Organizational Type to Certain Factors in Alabama's White Public Junior High Schools." (L.C. Card No. Mic 61-230) University of Alabama, (1960).

11. Dethy, Raymond Charles, Ph.D. "Relationships Between Educational Program Characteristics and Expenditure Level, Size and Kind of School Districts in Ohio." (L.C. Card No. 64-6893) Ohio State University, (1963).

12. Dungan, James Robert, Ed.D. "Relationships between Measures of Quality, Internally Variable Factors, and Externally Variable Factors in Selected Florida Secondary Schools." (L.C. Card No. 61-5519) University of Florida, (1961).

13. Evans, Byron F., Ed.D. "A Plan for the Reorganization of a Large High School into Several Coordinate Administrative Units." (L.C. Card No. 59-6413) University of Oregon, (1960).

14. Garcia, Genero Bruno, Ed.D. "Junior High School Size." (L.C. Card No. 61-3799) University of Southern California, (1961).

15. Gawkoski, Stephen Roman, Ph.D. "The Use of Community Characteristics for Obtaining Local Norms on Standardized Achievement Tests." (L.C. Card No. 12-434) Columbia University, (1955).

16. Grady, Charles E., Ed.D. "An Analysis of Administrative Practices of Large Six-Year High Schools of the North Central Association." (L.C. Card No. 20-568) University of Oklahoma, (1957).

17. Gray, Stuart Calvin, Ph.D. "A Study of the Relationship Between Size and a Number of Qualitative and Quantitative Factors of Education in Four Sizes of Secondary Schools in Iowa." (L.C. Card No. 61-5568) Iowa State University, (1961).

18. Hadden, Jeffrey K., "A Systematic Study of the City as a Unit of Analysis." (Ph.D. Dissertation, University of Wisconsin, (1963).

19. Hollitz, Erwin, Ed.D. "Factors Affecting the Establishment of a Second High School in a Community." (L.C. Card No. Mic 59-1425) Stanford University, (1959).

20. Holt, Charles C., Ed.D. "The Relationship of Large High School Internal Organization to Students and Staff as Individuals." (L.C. Card No. 60-3931) University of Illinois, (1960).

21. Indik, Bernard P., "Organization Size and Member Participation," Dissertation Abstract. 22 (1961), pp. 631-2.

22. Jackson, Joe LeRoy, Ed.D. "School Size and Other Factors Affecting Program Adequacy in Southern High Schools." (L.C. Card No. 66-4416) George Peabody College for Teachers, (1965).

23. Jackson, Jan Wayne, Ph.D. "Leadership, Size and Wealth as Related to Curricular Innovations in the Junior High School." (L.C. Card No. 66-6624) University of Michigan, (1965).

24. Jantz, Ralph Dale, Ed.D. "An Analysis of the Relationship of Accreditation, Finance, and Size of Nebraska High Schools to Scholastic Achievement." (L.C. Card No. 61-3850) University of Nebraska Teachers College, (1961).

25. Johnson, Glenn Hays, Ph.D. "A Search for Functional Unities: An Analysis of United States County Data." (L.C. Card No. 61-2585) New York University, (1958).

26. Kaplan, Howard Bernard, Ph.D. "An Empirical Typology for Urban Description." (L.C. Card No. 61-2628) New York University, (1958).

27. Kaufman, Walter Charles, Ph.D. "Social Area Analysis: An Explication of Theory, Methodology, and Techniques, with Statistical Tests of Revised Procedures, San Francisco and Chicago, 1950." (L.C. Card No. 62-851) Northwestern University, (1961).

28. King, Robert Edward, Jr., An Evaluation of the Decentralized Form of School Administration. Unpublished doctoral dissertation. George Washington University, (1966).

29. Koch, Edward Leo, Ed.D. "The Neighborhood School: The Relation of Social Area Factors and Residential Stability to Aggregate Pupil Achievement." (L.C. Card No. 66-5596) Cornell University, (1966).

30. Kovacs, Louis, Ed.D. "The Relationship of Selected Population Characteristics to Local Initiative in the Financial Support of Public Education in Virginia. (L.C. Card No. 66-179) University of Tennessee, (1965).

31. Lamberty, Earlyon John, Ed.D. "College Achievement in Relation to Size of High School from which Students Graduated." (L.C. Card No. Mic 65-7246) University of Nebraska Teachers College, (1965).

32. Lehman, James Louis, Ed.D. "Pupil Mobility and Its Relationship to Age, Intelligence Quotients and Achievement." (L.C. Card No. 64-2548) Northwestern University, (1963).

33. Lennon, Roger Thomas, Ph.D. "Prediction of Academic Achievement and Intelligence from Community and School-System Characteristics." (L.C. Card No. 4209) Columbia University, (1952).

34. McArthur, Harvey John, Ed.D. "A Survey of Academic Offerings and Courses Pursued by Academically Talented Pupils in Selected Kansas High Schools." (L.C. Card No. 63-802) University of Kansas, (1962).

35. Menozzi, John, Ed.D. "An Attempt to Determine the Optimum Size of Public Secondary Schools." University of Denver, (1959).

36. Oliver, Arnold D., Ph.D. "Relationship of Size of Connecticut School Districts and Certain Differences in Costs, Preparation of Staff, Services and Programs." (L.C. Card No. 66-879) University of Connecticut, (1965).

37. Osburn, Morris, Ed.D. "The Effect of Expenditure per Pupil and Size of School on the Quality of Education in the Secondary Schools in Mississippi." (L.C. Card No. 63-1777) University of Southern Mississippi, (1962).

38. Patterson, Harold Dean, Ed.D. "Relationship Between Size of Secondary School and Selected Teacher Characteristics." (L.C. Card No. Mic 65-3555) George Peabody College for Teachers, (1964).

39. Pierce, Keith Allyn, Ed.D. "Dropouts, Delinquents, and Their Student Peers." (L.C. Card No. 65-12,231) University of Oregon, (1965).

40. Rosenthal, Oscar Arthur, Ph.D. "A Semantic Differential Investigation of Critical Factors Related to Achievement and Underachievement of High School Students." (L.C. Card No. 65-12,265) University of Southern California, (1965).

41. Shapiro, David Franklin, Ed.D. "Relationship of High School Size to Staff Relations." (L.C. Card No. 58-1277) Stanford Univessity, (1958).

42. Shelley, Herman Walker, Ed.D. "An Analysis of the Relationships Between Eight Factors and Three Measures of Quality in Thirty-nine South Carolina Secondary Schools." (L.C. Card No. 20-770) University of Florida, (1957).

43. Smith, Clifford Basil, Ph.D. "A Study of the Optimum Size of Secondary Schools." (L.C. Card No. Mic 60-6408) Ohio State University, (1960).

44. Thomas, James Alan, Ph.D. "Efficiency in Education: A Study of the Relationship Between Selected Inputs and Mean Test Scores in a Sample of Senior High Schools." (L.C. Card No. 63-2747) Stanford University, (1962).

45. Treadway, John F., Ed.D. "A Study of Certain Factors Associated with the Quality of Education in Arkansas School Districts." (L.C. Card No. 62-2285) University of Arkansas, (1962).

46. Tyson, James Carroll, Ed.D. "A Comparative Study of Teacher-Pupil Relationship in Small and Large High Schools." (L.C. Card No. 22,909) University of Virginia, (1957).

47. Vardeman, Martha Hay, Ph.D. "A Study of the Relationship Between Size and Accreditation of Schools and Certain Aspects of the Instructional Program in Public Negro High Schools, Alabama, 1958." (L.C. Card No. Mic 60-1729) University of Alabama, (1959).

48. Watson, Barclay Herbert, Ed.D. "A Study of the Relationship of School Size and Grade Level to Administrative Problems." (L.C. Card No. 64-4218) Colorado State College, (1963).

49. Weaver, Charles Horace, Ph.D. "An Investigation of the Influence of Size on the Quality of the High School." (L.C. Card No. 62-3156) University of North Carolina, (1961).

50. Williams, Robert Hattler, Ed.D. "Some Characteristics of Washington State University Students as Related to the Size of High School they Attended." (L.C. Card No. Mic 65-1539) Washington State University, (1964).

51. Young, Nathan, Ed.D. "Community Predictors of School Holding Power." (L.C. Card No. 63-7557) Yeshiva University, (1963).

52. Zinser, Ernest Melvin, Ed.D. "Teacher Morale and Size of Secondary Schools." (L.C. Card No. Mic 67-3211) Arizona State University, (1967).

APPENDIX

APPENDIX A

SCHOOL AND COMMUNITY UNITS*

While other units of local government, particularly the special purpose districts, have increased steadily in numbers, the local school district has decreased phenomenally. Since 1932, the total number of school districts has declined from 127,649 to just above 25,000. The decline is due almost exclusively to the reorganization of rural school districts, most of which were operating a one-teacher elementary school or no school at all. In the intervening 35 years the population of the U.S. has increased more than 50%. The number of local governmental units in 1962 by type for the U.S. and for Iowa and Florida are given below:

	U. S.	Iowa	Florida
counties	3043	99	67
municipalities	18000	944	366
townships	17142	-	-
special districts, exclusive of school districts	18323	263	264
school districts, exclusive of 2341 public school systems operated as part of other governmental units	34678	1336	67
total	91186	2642	764

Most of the decline since 1952 in numbers of school districts has been confined to ten states: Illinois, Iowa, Kansas, Michigan, Minnesota, Mississippi, Missouri, Nebraska, New York, and Wisconsin. Moreover, most of the change was legislated at the state level.

As of 1960, the average area in square miles of school districts in middle western states was under 50. It was over 200 in southern states, and over 1000 in western states.

The relationships of school district boundaries to those of other units of local government are quite varied. In three states (Florida, Nevada, and West Virginia) the independent school units are coterminous

---

*U.S. Department of Commerce, Statistical Abstract of the U.S., 89th Edition. Washington, D.C.: Government Printing Office, 1966. pp. 413-416.

with counties. In five others (Alabama, Georgia, Kentucky, Louisiana and Utah) the independent school unit boundaries are generally but not universally coterminous with counties. In four states (Indiana, New Hampshire, New Jersey, and Pennsylvania) school units are coterminous with townships or towns. In all others, there is only occasionally a coincidental relationship with other units of local government. Even the earlier identification of school districts with municipalities has been largely lost.

Education may be described as a function of population. Consequently, density of population controls to a considerable extent the geographic spread of school units. For example, one school district in Nevada (Elko County) exceeds by almost 3000 square miles the geographic area of 562 school units in Connecticut, Massachusetts, and Rhode Island combined. The relative enrollments are 2900 and 1.2 million.

States may be classified according to the numbers of different types of school units found in each, as follows:

one type only	4
two types	13
three types	15
four types	6
five types	7
six types	3
eight types	2

In some states, the classification is based on governmental units such as municipalities, towns, and counties. In others, it is based at least in part on units of population such as total population of the district or school census. In yet others, school grade organization is used to distinguish units (e.g., elementary, high school, junior college). In most states, as shown above, no single method is used.

The definitions given below are useful in bringing the problem of mapping school districts in terms of census into proper perspective. The definitions are divided into two sets, one clarifying the school units, and the other, the census units. In general, the two sets of units are not geographically coterminous.

Definitions of School Units

A. SCHOOL DISTRICT: A school district is generally a local unit of government, possessing quasi corporate powers established or empowered by state law to conduct and administer a public school or a system of public schools. The term includes all types of school districts, such as common, city, independent consolidated, separate high school, union, community township, and county unit school districts. The term "administrative unit" is often used.

1. A COMMON SCHOOL DISTRICT is an autonomous local subdivision created only for school purposes and operated by a board or school official. It usually conducts only one school.

2. A COMMUNITY SCHOOL DISTRICT is of the same nature as a common school district. Its distinguishing characteristic is that it comprises the territory of a whole community usually composed of a village, town, or city and its tributary trade and service area of smaller population centers and open country.

3. A CITY OR INDEPENDENT SCHOOL DISTRICT is one that comprises at least the area of an entire city. It usually is a separate quasi corporation independent of the general municipal government.

4. A TOWN OR TOWNSHIP SCHOOL DISTRICT is one that is coterminous with a political town of a New England state, or with a political township as in Indiana, Pennsylvania and New Jersey. Such districts are usually independent of the town or township government.

5. A COUNTY SCHOOL DISTRICT-COUNTY UNIT is one that is coterminous with a civil county. Such districts are not a part of general county government and are usually largely or wholly autonomous in the administration of schools. Such districts are of two types:

   a. A COMPLETE COUNTY UNIT is a district which comprises an entire county; and
   b. A PARTIAL COUNTY UNIT is a district that comprises all of a county outside of one or more city or independent districts.

6. A SEPARATE HIGH SCHOOL DISTRICT is one responsible only for administering a high school. Such a district is superimposed upon the territory of one or more elementary school districts and is usually not coterminous with any other political subdivision. In some instances, however, it is coterminous with a township as in Illinois and Wisconsin. The term HIGH SCHOOL includes any type: the traditional 4-year high school, 2-year or 3-year senior high schools, or 6-year or 5-year high schools.

B. ATTENDANCE UNITS: An attendance unit or school unit comprises the geographical area and its population served by a single school. In larger districts it is part of a school district.

Definitions of Census Units

A. GENERAL

1. URBAN-RURAL--An urban population comprises all persons living in (a) places of 2,500 inhabitants or more incorporated as cities, boroughs, villages and towns; (b) the densely settled urban fringe, whether incorporated or unincorporated, of urban areas; (c) unincorporated places of 2,500 inhabitants or more.

2. PLACES--Concentrations of population, regardless of the existence of legally prescribed limits, powers, or functions. Most of the places listed are incorporated as cities, towns, villages, or boroughs, however. The larger unincorporated place outside the urbanized area was delineated and those with a population of 1,000 or more are presented in the same manner as incorporated places of equal size.

3. ENUMERATION DISTRICT--A small area assigned to an enumerator which must be canvassed and reported separately. In most cases an ED contains approximately 250 housing units. The boundaries for the ED's for the most part follow such features as roads, streets, railroads, streams, and other clearly defined lines which may be easily identified by census enumerators in the field and often do not conform to the boundaries of political units.

B. SPECIFIC

1. COUNTIES--The primary divisions of the States except for Louisiana where the divisions are called parishes and in Alaska where the data are collected for election districts.

2. MINOR CIVIL DIVISIONS--The primary political divisions into which counties are divided. Where more than one type of primary division exists in a county, the bureau uses the more stable division so that comparable data are available from census to census (School, taxation, election units, etc. are not considered stable.).

3. STANDARD METROPOLITAN STATISTICAL AREAS (SMSA)--A unit whose entire population is in and around the city whose activities form an integrated social and economic system. Except in New England, an SMSA is a county or group of contiguous counties which contain at least one city of 50,000 inhabitants or more or "twin cities" with a combined population of at least 50,000. In addition to the counties containing such a city or cities, contiguous counties are included in an SMSA, if, according to certain criteria they are essentially metropolitan in character and are socially and economically integrated with the central city. Criteria are listed on page viii of PC(1).

4. URBAN AREAS--An urbanized area contains at least one city of 50,000 inhabitants or more in 1960, as well as the surrounding closely settled incorporated places and unincorporated areas that meet the criteria listed on page vii of BULLETIN PC(1).

5. URBAN PLACES--All incorporated and unincorporated places of 2,500 inhabitants or more, and the towns, townships, and counties classified as urban.

6. INCORPORATED PLACES--Places which are incorporated as cities boroughs, towns, and villages with the exception that towns are not recognized as incorporated places in the New England states, New York and Wisconsin.

7. UNINCORPORATED PLACES--Densely settled population centers without corporate limits.

Census information is divided in three different ways as follows:

A. MAJOR SUBJECT FIELDS
   1. Population
   2. Housing
   3. Agriculture
   4. Business
   5. Manufactures
   6. Mineral Industries
   7. Transportation
   8. Governments

B. GEOGRAPHIC AREAS
   1. United States
   2. Regions and Divisions
   3. States
   4. Counties
   5. Standard Metropolitan Statistical Areas
   6. Minor Civil Divisions
   7. Cities and Other Incorporated and Unincorporated Places
   8. Urbanized Areas
   9. Census Tracts

C. RETAIL TRADE AREAS
   1. Central Business Districts
   2. Major Retail Centers
   3. City Blocks
   4. Enumeration Districts
   5. Foreign Trade Statistical Areas

APPENDIX B

ILLUSTRATION OF THE PROCESS OF CONTENT ANALYSIS

The following page of this appendix contains the file documentation for Iowa File DB-16. This foldout traces a completed data item file sheet for one item -- "semester hours". The steps in completing the file sheet were:

a) The unit which the item describes--teacher--was inferred from the file identification page since the name of the file is Teacher Employment. The unit was recorded on the data item file sheet and applied to all items in the file.

b) The item "Sem Hrs" was located as the next item to be processed on the information layout form. The abbreviated title of the item was copied on the file sheet. (All other items in File DB-16 were also first located on this form.)

c) The item was located on the source form. The coding system and the source used in its determination were recorded. The meaning of the item was generally clear -- a teacher's total semester hours of college credit earned.

d) A search was made for a more precise description of the item. The information was found on page 5 of the Iowa Educational Data Bank User's Manual. It was assumed that the item in the manual and the item being processed were the same even though the manual described the item as having no coding table. A key-word title was constructed and recorded according to the criteria presented in Chapter 4 Section B. The complete description and source were recorded.

e) The physical information was recorded from the file identification page.

INFORMATION LAYOUT FORM

information for this layout ... punch cards supplied by Evel... at the State Department

1 59-60	FOLDER NUMBER	TITLE, POSITION, EXP. THIS, TOTAL, CERT. AND YEAR, EMP. YR., CO., SCHOOL DISTRICT, SCHOOL ... SAL.
2 60-61 and 61-62	FOLDER NUMBER	NAME, TITLE, AGE, COUNTY NO.
3	See Source for	
4		
5		
6		

1. Tape is card to tape (Block 10, Record le...
2. All three years are loaded separately in ...
3. Note information in 59-60 file is incompl...

MULTIPLE CARD FO...

DATA ITEM FILE SHEET

Iowa Item No. 0135

Unit described by ite...

Abbreviated title of i...

KW title: Teachers Total Semester Hours of College ...

Teachers Preparation

Further Description

Total semester hours of college ... undergrad. and grad. level. (Quarter-hrs of credit converte... equivalent by multiplying quart...

Source (s) for description: Ia. Ed. Data, p. 5, ...

Coding system:

Code	Number of Hrs.
0	None
1	1 - 29
2	30 - 59
3	60 - 89
4	90 - 119
5	120 - 149
6	150 - 179
7	180 - 209
8	210 or above

Source (s) for coding system: DB - 16 Source

...ical information:

File No. 016 File Name ...

Iowa Format Book, page 25.

CO... IT...

DATA BANK

File # 016

Teacher Employment 59-62

File Name

Date Filed at MRC

Tape Unit ...
Blocking factor
Record length
Physical size
... Size

File First Created:
Date:
By:
Subsequent ...

FILE IDENTIFYING PAGE

SOURCE FORM

D B - 16 Source form

000

SCHOOL:

4	5	6	7	8	9	10
AGE LAST BIRTHDAY	COUNTY NUMBER	SCHOOL DISTRICT NUMBER	POSITION (See Code below)	YEARS EXPERIENCE: THIS SCHOOL DISTRICT	YEARS EXPERIENCE: TOTAL	HIGHEST DEGREE HELD

MR. MISS. MRS.

NONE / BACHELOR / MASTER'S / SPECIALIST / DOCTOR'S

TOTAL SEMESTER HOURS EARNED

NONE 0; 1-29 1; 30-59 2; 60-89 3; 90-119 4; 120-149 5; 150-179 6; 180-209 7; 210 OR ABOVE 8

...tem teacher

...f item Sem Hrs

...ge Credit --

...ge credit earned at

...ted to semester-hrs.
...rter-hrs by 2/3.)

415

...e Iowa, I I B p. 26

Teacher Employment 59-62

COMPLETED IOWA DATA
...TEM FILE SHEET

Reference Number	Item	Table
	This District - Total consecutive years individual has been employed by the district in which they are currently employed. Current year is not counted.	
	Total - Total years active professional experience.	
110.	Contract Periods - Number of days individuals are required to serve under the terms of the current contract.	N/A
111.	Percent of Time - The full-time equivalency of part-time personnel expressed as a percentage.	N/A
112.	Salary - The total professional contract salary, including pay for extra duties and other compensation which may be in addition to base salary.	N/A
113.	Extra Compensation - A code indicating reason or reasons for individuals' salary exceeding the normal step on the salary schedule.	3
114.	Occupation Previous Year - A code indicating individuals' occupation during previous school year.	4
115.	Total Semester Hours - Total Semester hours of college credit earned at undergraduate and graduate level. (Quarter-hours of credit converted to semester hours equivalent by multiplying quarter hours by 2/3).	N/A
117.		
118.	Declared Majors - Graduate - One or more (maximum of 4) codes indicating areas or subject matter fields in which the individual took a major(s) at the undergraduate level.	6
119.	Vocational Years Work Experience - The number of years of commercial work experience	N/A

PAGE FROM INFORMATION MANUAL

APPENDIX C

THE MAPPING PROCEDURES

Preliminary Tests

The Iowa State Department of Public Instruction has prepared a map of the state with all secondary school districts indicated in a color code. The state map was constructed from a variety of county maps of more or less precision. These were photographed, and 35mm slides were prepared. Efforts to superimpose a minor civil division map on the school district map were fruitless. The University of Wisconsin Geography Department advised that the time and money needed to achieve reasonable accuracy were prohibitive.

Professor Randall Sale experimented with estimation techniques to determine if estimation time could be reduced. He found that after some practice, he could obtain proportional estimates not significantly different from those obtained by using a planimeter.

Another device sometimes useful was counting sections within townships. Not all the county maps displayed these sections, but for those which did, percentages were easily and accurately calculated.

The remaining pages of this appendix comprise an instruction booklet which aided in carrying out the mapping task.

INSTRUCTION BOOKLET

A. Problem:

Assign each township or part of township, town or part of town, census tract or part of census tract to a specific school district.

B. Materials Needed:

1. State map with all school districts identified by code numbers.
2. All county maps with school district lines clearly indicated.
3. Maps for all SMSA's in the state.
4. A print-out of the list of counties, townships, towns, etc. of the states with four dashes and slashes dividing the space to the right of the name.
   For example: Blair Twp. ____-____/____-____/____-____/____-____
5. A table of proportional parts enabling workers to convert numbers quickly to percent of the township. (Most townships had 36 sections--see Appendix C, Figure 1.)
6. A list of all school districts both alphabetically and numerically arranged for cross-checking purposes. This listing should be by county, the counties being alphabetically arranged. (See Appendix C, Figure 2.)

C. Personnel Needed:

1. Staff instructor to train and supervise local people.
2. Three pairs of workers; the first of each pair makes estimates of area and the second writes percents on the print-out.

D. Process:

1. Divide the state into three parts geographically so that each pair of workers is responsible for approximately the same number of counties.
2. Display the large state map indicating all minor civil divisions and explain how the state is divided into townships and the townships into sections.
3. Practice estimating areas on townships (see Appendix C, Figure 3), and then cross check the results using the table of proportional parts. For example, 12 sections of the 36 section township would be equivalent to 33% of the township.
4. After a couple of practice estimates verified by the staff instructor, supply each pair of workers with a state map on which the school districts are numerically listed and a county map with the school district lines clearly drawn and the various school district names indicated.
5. Have one worker do the estimating of area and the other enter the estimates (percents) on the census print-out. Percents are listed after the number of the school district. For example:

   3461-70%/329-20%/3817-10%/___-___

No. of Sections	%	No. of Sections	%
1	3	19	53
2	6	20	56
3	8	21	58
4	11	22	61
5	14	23	64
6	17	24	67
7	19	25	69
8	22	26	72
9	25	27	75
10	28	28	78
11	31	29	81
12	33	30	83
13	36	31	86
14	39	32	89
15	42	33	92
16	44	34	94
17	47	35	97
18	50	36	100

Appendix C, Figure 1

Table of Porportional Parts

24 1845 Dow City-Arion Comm Sch Dist
24 3996 Manilla Comm School Dist
24 5832 Schleswig Comm Sch Dist

Delaware
28 1989 Edgewood Colesburg Comm Sch Dist
28 4043 Maquoketa Valley Comm Sch Dist
28 6950 West Delaware Comm Sch Dist
28 9025 Bremen Twp. Sch Dist

Dickinson
30 342 Arnolds Park Cons Sch Dist
30 2846 Harris Lake Park Comm Sch Dist
30 4284 Milford Comm School Dist
30 6120 Spirit Lake Comm Sch Dist
30 6345 Terril Comm School Dist

Guthrie
39 522 Bayard Community Sch Dist

Kossuth
55 126 Algona Comm School Dist
55 900 Burt Community Sch Dist
55 3456 Lakota Cons School Dist
55 3573 Ledyard Comm School Dist
55 3897 Lu Verne Comm Sch Dist
55 5868 Sentral Comm School Dist
55 6309 Swea City Comm Sch Dist
55 6417 Titonka Cons School Dist
55 9022 Bancroft Cons School Dist
55 9026 Greenwood Twp Sch Dist
55 9041 Ramsey Rur Ind Sch Dist
55 9047 Grant Cons School Dist

14 9057 Boise Rur Ind Sch Dist
14 9058 Lincoln Rur Ind Sch Dist
14 9059 Granville Rur Ind Sch Dist
14 9060 Sheridan Rur Ind Sch Dist
14 9062 Buck Run Rur Ind Sch Dist
14 9063 Prairieville Rur Ind S D
14 9064 Storm Creek Rur Ind S D

Dubuque
31 1863 Dubuque Comm School Dist
31 6961 Western Dubuque Comm Sch Dist

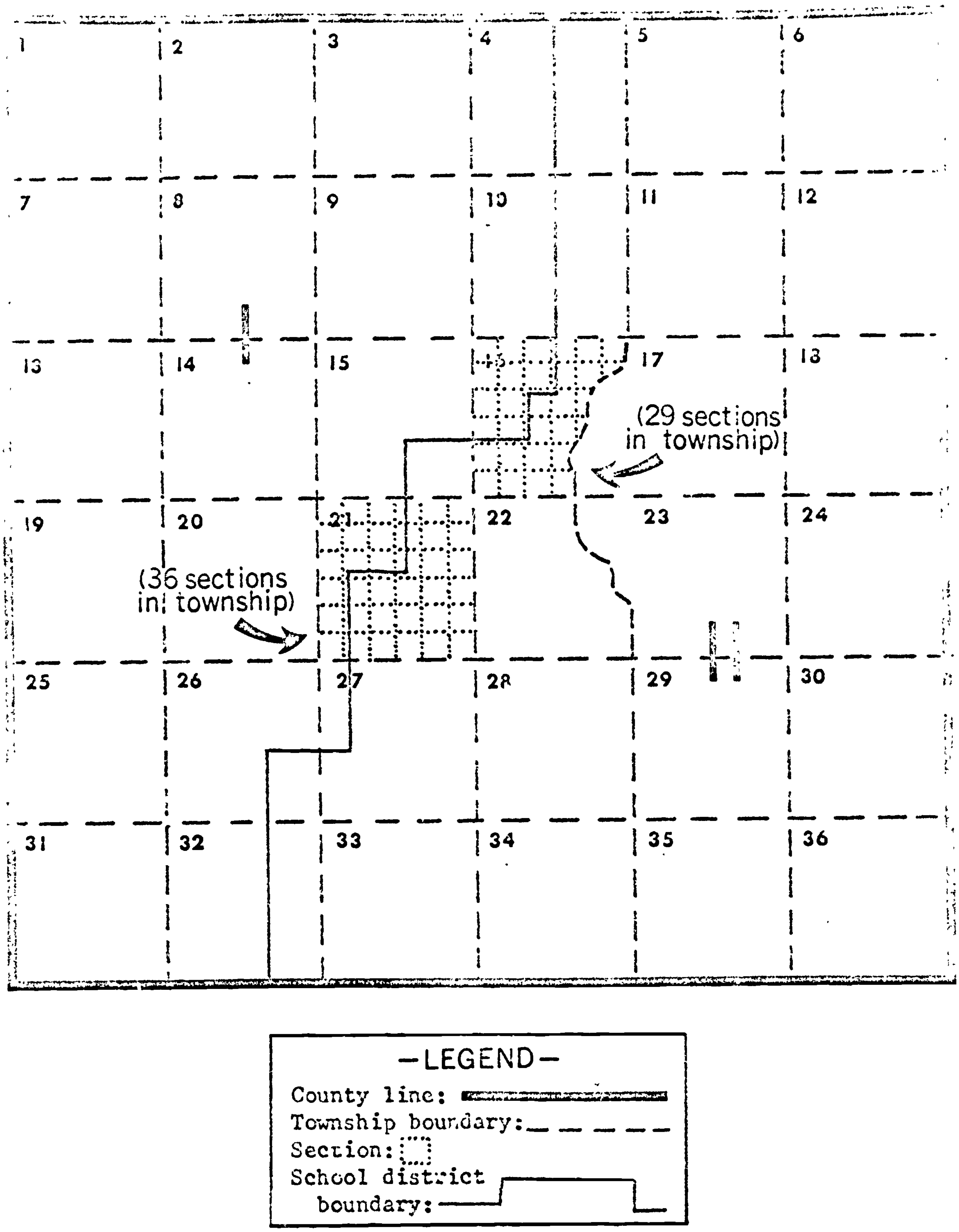
Emmet
32 333 Armstrong Comm Sch Dist
32 2144 Estherville Comm Sch Dist
32 3700 Lincoln Central Comm Sch Dist
32 5544 Ringsted Comm School Dist

Jackson
49 243 Andrew Comm School Dist
49 585 Bellevue Comm Sch Dist
49 4041 Maquoketa Comm Sch Dist
49 4275 Miles Comm School Dist
49 5337 Preston Comm School Dist
49 5733 Sabula Comm School Dist

Palo Alto
74 450 Ayrshire Cons Sch Dist
74 2088 Emmetsburg Comm Sch Dist (Cylinder Joined)
74 2556 Graettinger Comm Sch Dist
74 3969 Mallard Comm School Dist
74 5724 Ruthven Cons School Dist
74 6921 West Bend Comm Sch Dist

Appendix C, Figure 2

List of School Districts



APPENDIX C, FIGURE 3
Practice for areal estimating,
superimposing map elements upon a hypothetical county.

Some parts of Iowa townships were not joined to a secondary school district at the time of the study so that the total percent was not always 100%.

6. Have each pair proceed county by county until they complete their geographical area.
7. SMSA's were divided by percent of census tract.

Transformation of the Census Data

Except for Case IV (see Table 4-C-2), the proportional correspondences of the disconsonant units had been determined through mapping of area. Enough information had been gathered to verify that a particular district derives its population from 25% of one township plus 100% of another, etc. The single district in Case IV had two high schools. Each attendance area received an undifferentiated allocation of census data in proportion to its area.

All the information was present. In fact, more information had been gathered than could be used. Basically, two lists were present: the first was a list of actual census data units, units for which data were stored on magnetic tape; the second was a list of all the geographic areas in Iowa, areas which had been assigned to school districts.

APPENDIX D

INDICATOR DEFINITIONS, INTERCORRELATIONS, AND MEANS: THE 48 STATE SURVEY

A number of sources were used to obtain data for the 48 state survey. Demographic and general educational data were obtained from Statistical Abstract of the U.S. and to a lesser extent from the U.S. Book of Facts Statistics and Information. For more detailed educational data two types of references were used. The first included two government publications on educational statistics, Public Secondary Schools Number 1 and the Digest of Educational Statistics. The second was Ranking of the States published by the National Educational Association. Other references were used to a more minor extent. Below are listed the sources used.

Construct Teacher Quality

Indicators	Definition	Source
1. Average secondary school teacher's salary.	The average salary per public secondary school teacher by state, 1961.	SA, 1962, p. 129
2. Percent of male teachers.	Percent of public secondary school teachers who were men by state, 1959.	PSS, 1961, p. 44-5
3. Supervisors per pupil.	The number of department heads and supervisors per public secondary school pupil multiplied by 1000 per state, 1959.	PSS, 1961, p. 44-5
4. Guidance counselors per pupil.	The number of public secondary school guidance counselors per pupil multiplied by 1000 per state, 1959.	PSS, 1961, p. 44-5
5. Pupil teacher ratio.	The number of public secondary school pupils per public secondary school teacher per state, 1959.	PSS, 1961, p. 16
6. Librarians per secondary school pupil.	The number of public secondary school librarians per secondary school pupil multiplied by 1000 per state, 1959.	PSS, 1961, p. 44-5

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
5175.062500	770.807677	5243.00	5025.00	+84
52.543750	7.857160	58.30	47.80	+62
.707292	.505296	0.34	0.79	+76
1.570208	.553521	1.49	0.98	+84
22.266667	2.362761	18.80	24.40	-42
1.545625	.409191	1.31	1.40	-64

Intercorrelations

	1	2	3	4	5	6
1	100	61	55	54	-16	-49
2	61	100	11	44	-44	-10
3	55	11	100	62	-13	-62
4	54	44	62	100	-46	-41
5	-16	-44	-13	-46	100	-11
6	-49	-10	-62	-41	-11	100

Construct School Size

	Indicators	Definition	Source
1.	Percent enrollment in small secondary schools.	The percent of public secondary school pupils enrolled in schools of 1-499 pupils by state, 1959.	PSS, 1961, p.12-13
2.	Percent enrollment in medium sized secondary schools.	The percent of public secondary school pupils enrolled in schools of 500-1499 pupils by state, 1959.	PSS, 1961, p.12-13
3.	Change in number of secondary schools.	The change in the number of public secondary schools 1950 to 1960 expressed as the ratio of the 1960 figure to the 1950 figure per state.	SA, 1953, p. 119,
4.	Additional classrooms needed.	Number of additional classrooms needed to reduce the public secondary school class size to 30 secondary school pupils. This figure is divided by total public secondary school pupils and multiplied by 1000 per state, 1962.	DES, 1964, p. 40,
5.	Excess public secondary school pupils.	The number of public secondary school pupils in excess of normal plant capacity divided by the total public secondary school pupils and multiplied by 100 per state, 1961.	SA, 1962, p. 124

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
45.664583	18.190101	63.00	26.10	-90
38.872917	11.222387	32.10	45.10	+89
1.069167	.213472	0.73	1.11	+68
4.306875	2.086977	1.99	7.17	-11
6.052083	3.863126	3.50	9.10	+03

Intercorrelations

	1	2	3	4	5
1	100	-76	-44	-06	-01
2	-76	100	37	-18	01
3	-44	37	100	-03	04
4	-06	-18	-03	100	13
5	-01	01	04	13	100

Construct State Control

Indicators	Definition	Source
1. Number of state-required courses.	Number of new courses (as of Dec. 1955) required with content and methods subject to state or local authority.	CR, 1960, p. 13
2. Number of state-recommended courses.	Number of new courses (as of Dec., 1955) recommended by state or local authorities.	CR, 1960, p. 13
3. Number of state-guided courses.	Number of new courses (as of Dec., 1955) either required or recommended by state authorities.	CR, 1960, p. 13

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
2.604167	2.497829	0.0	4.0	+62
13.958333	6.058597	9.0	19.0	+79
6.479167	4.949704	0.0	17.0	+82

Intercorrelations

1	100	23	29
2	23	100	48
3	29	48	100

Construct Funding

Indicators	Definition	Source
1. Percent state funds.	Percent of revenue receipts for public elementary and secondary schools from state sources per state, 1961-2.	DES, 1964, p. 58-9
2. Percent federal funds.	Percent of revenue receipts for public elementary and secondary schools from federal sources per state, 1961-2.	DES, 1964, p. 58-9
3. Per-capita state educational expenditure.	The total state expenditure on lower education divided by the total state population, 1960.	SA, 1962, p. 12, p. 113
4. Percent spent on education.	Direct general expenditure of state and local government by function-percent spent on education per state, 1960.	SA, 1962, p. 423
5. Percent spent on public welfare.	Direct general expenditure of state and local government by function-percent spent on public welfare per state, 1960.	SA, 1962, p. 423
6. State educational expenditure per pupil.	The total state expenditure on lower education divided by the total public school pupils (elementary and secondary) per state, 1960.	SA, 1962, p. 113

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
39.537500	18.191042	11.30	52.20	-42
5.608333	2.207264	3.50	5.70	-70
69.958750	13.935488	77.17	58.10	83
36.845833	4.136220	37.00	32.60	08
8.522917	3.104247	8.20	6.30	-45
365.083333	83.343624	400.00	310.00	93

Intercorrelations

	1	2	3	4	5	6
1	100	26	-14	26	07	-32
2	26	100	-32	09	18	-59
3	-14	-32	100	31	-28	81
4	26	09	31	100	-20	-06
5	07	18	-28	-20	100	-25
6	-32	-59	81	-06	-25	100

Construct Quantitative Population Change

	Indicators	Definition	Source
1.	Population increase.	Percent change in U.S. population by state, 1950-60.	USB, 1966, p. 13
2.	Population mobility.	Percent of 1960 population who resided in a different county (or abroad) from their 1955 residence.	USB, 1966, p. 33
3.	Birthrate.	The number of live births per 1000 population for each state, 1960.	SA, 1961, p. 50
4.	Population under 15 years.	The number in thousands of persons under 15 years per state, 1960.	USB, 1966, p. 23
5.	Change in public secondary school pupils.	The 1960 figure for public secondary school pupils divided by the 1950 figure per state.	SA, 1953, p. 120, 1962, p. 125
6.	Percent of population under 15 years old.	The percent of the population under 15 years by state, 1960.	USB, 1966, p. 11, p. 23

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
18.910417	18.731983	5.20	78.70	+84
17.681250	5.638139	14.50	31.00	+82
24.408333	2.180485	23.30	23.30	+70
1151.916667	1106.634516	858.00	1467.00	-22
1.510417	.288241	1.24	2.31	+85
31.906250	2.239829	31.10	29.30	+48

Intercorrelations

1	100	73	32	09	90	01
2	73	100	35	-20	66	15
3	32	35	100	-28	35	85
4	09	-20	-28	100	01	-29
5	90	66	35	01	100	12
6	01	15	85	-29	12	100

Construct Qualitative Population Change

Indicators	Definition	Source
1. Net migration white population.	The migration from abroad, interdivisional, interregional, and interstate of U.S. white population, 1950-60.	USB, p. 34
2. Net migration non-white population.	The migration from abroad, interdivisional, interregional, and interstate of U.S. non-white population, 1950-60.	USB, 1966, p. 34
3. Percent Negro.	The percent of U.S. population who are Negro, by state, 1960.	USB, 1966, p. 27
4. Change in urbanization.	The change in urbanization between 1950 and 1960 per state.	USB, 1966, p. 16
5. Percent urbanization.	The percent of the state's population living in urban areas (2500 or more people), 1960.	USB, 1966, p. 16
6. Percent engaged in manufacturing.	Percent of employees in non-agricultural establishments working in manufacturing, 1960.	SA, 1962, p. 224
7. Percent of technical and professional workers.	The percent of the state's population who were technical or professional workers, 1960.	RS, 1966, p. 59

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
2.139583	17.315982	-9.10	70.00	+39
12.566667	32.875823	12.30	15.60	+81
8.910417	10.479330	0.90	17.80	-62
7.008333	5.636408	5.40	8.50	-51
60.993750	15.417663	53.00	73.90	+80
28.531250	14.292352	26.00	15.60	-08
11.304167	3.810400	9.70	10.30	-09

Intercorrelations

1	100	32	01	29	44	-07	25
2	32	100	-45	-33	46	25	14
3	01	-45	100	19	-28	26	-11
4	29	-33	19	100	-31	01	60
5	44	46	-28	-31	100	-08	-04
6	-07	25	26	01	-08	100	49
7	25	14	-11	60	-04	49	100

Construct Social Responsibility

Indicators	Definition	
1. Percent voting.	Percent of voting age population who voted for president in 1960 by state.	USB, 1966, p. 381
2. Percent covered by hospital insurance.	Percent of 1960 population covered by hospital insurance by state.	SA, 1962, p. 479
3. Expenditure for parks.	The current operating expenditure for municipal and county parks 1960 calculated on a per capita basis by state, 1960.	SA, 1962, p. 12, p. 205
4. Amount raised by community chest.	The amount raised by community chest and united fund campaigns 1960 calculated on a per capita basis by state, 1960.	SA, 1962, p. 12, p. 306
5. Membership in AFL-CIO.	The membership in AFL-CIO per state calculated on a per capita basis and multiplied by 100, 1960.	SA, 1962, p. 12, 1963, p. 249

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
64.093750	15.322919	76.50	50.00	+64
68.502083	11.837396	59.80	66.80	+89
1.582083	.975923	0.99	2.74	+73
2.0[illegible]8125	.967612	2.12	1.80	+78
5.989583	3.205804	4.89	3.00	+87

Intercorrelations

1	100	58	29	27	44
2	58	100	49	69	66
3	29	49	100	42	66
4	27	69	42	100	60
5	44	66	66	60	100

Construct Educational Attainment

Indicators	Definition	Source
1. Percent with 4 years of high school.	The number of individuals 25 years old and older who have completed 4 or more years of high school 1960 divided by the total state population 25 years and over, multiplied by 100, 1960.	DES, 1964, p. 125
2. Percent with 4 years of college.	The number of individuals 25 years old and older who have completed 4 or more years of college 1960 divided by the total state population 25 years and over, multiplied by 100, 1960.	DES, 1964 p. 125
3. Median school years.	The median number of school years completed by individuals 25 years and over, 1960.	RS, 1966 p. 25
4. Percent illiterate.	The percent of each state's population 14 years old and older who were illiterate, 1960.	RS, 1966, p. 25
5. Number of doctors.	The number of doctors per state per 100,000 population, 1960.	SA, 1962, p. 74
6. Number of psychologists.	The number of psychologists per state per 100,000 population, 1960.	APA, 1961, pp. 806-943
7. Percent of professional and technical workers.	The percent of the state's population who were technical or professional workers, 1960.	RS, 1966, p. 59
8. High school graduates.	Public high school graduates 1959 as a percent of 1955-6 ninth grade enrollment.	SSS, 1959, pp. 56-7

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
41.170833	7.165047	46.30	30.60	+89
7.443750	1.518484	6.40	7.40	+82
10.552083	1.093540	11.30	10.90	+90
2.252083	1.386541	0.70	2.60	69
119.250000	28.442559	117.00	131.00	+71
8.035417	4.258300	7.70	7.60	+77
10.856250	1.649609	9.70	10.30	+85
67.012500	6.846947	76.50	65.60	+59

Intercorrelations

1	100	70	96	-72	41	44	68	59
2	70	100	71	-23	56	66	90	19
3	96	71	100	-71	47	49	69	55
4	-72	-23	-71	100	-33	-34	-35	-74
5	41	56	47	-33	100	86	60	24
6	44	66	49	-34	86	100	72	31
7	68	90	69	-35	60	72	100	25
8	59	19	55	-74	24	31	25	100

Construct Ability to Pay

Indicators	Definition	Source
1. Per-capita income.	The per-capita income 1960 determined for each state	SA, 1962, p. 319
2. Change in per-capita income.	The percent change in per-capita income between 1950 and 1960 per state.	SA, 1962, p. 322
3. Poverty.	Percent of population with income under $2000 per state, 1959.	SA, 1962, p. 333
4. Percent of non-public secondary schools.	Percent of total **school** enrollment 1960 who are enrolled in non-public schools per state.	SA, 1963, p. 128, p. 131
5. Change in non-public secondary school	The percent of all secondary school pupils attending non-public schools 1960 divided by the comparable figure for 1950 per state.	SA, 1953, p. 129, p. 133, 1963, p. 128, p. 131

Means	Standard Deviations	Iowa Value	Florida Value	Principal Component Loading
2064.645833	423.093542	2003.0	1988.0	+92
77.854167	29.238238	46.0	174.0	+24
14.641667	7.359966	15.2	16.2	-93
10.172917	6.402107	11.0	10.5	+71
1.127083	.272748	1.1	1.9	-08

Intercorrelations

1	100	26	-86	46	04
2	26	100	-19	-10	42
3	-86	-19	100	-53	-04
4	46	-10	-53	100	-40
5	04	42	-04	-40	100

OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

ERIC ACCESSION NO.				
CLEARINGHOUSE ACCESSION NUMBER	RESUME DATE 15 04 68	P.A.	T.A.	IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒ / ERIC REPRODUCTION RELEASE? YES ☒ NO ☐

TITLE

SIZE FACTORS AND NON-DOLLAR COSTS OF SECONDARY SCHOOLS, PHASE 1, FINAL REPORT, PROJECT NO. 6-1266, 6/1/66-8/31/67.

PERSONAL AUTHOR(S)

WAKEFIELD, HOWARD E. * AND OTHERS

INSTITUTION (SOURCE)

SOURCE CODE ZQU98400

REPORT/SERIES NO.

OTHER SOURCE

SOURCE CODE

OTHER REPORT NO.

OTHER SOURCE

SOURCE CODE

OTHER REPORT NO.

PUB'L. DATE 15 – APR – 68

CONTRACT/GRANT NUMBER OEC 3-6-061266-0801

PAGINATION, ETC.

216p.

RETRIEVAL TERMS

IDENTIFIERS

ABSTRACT

ELEMENTS OF EDUCATIONAL PRODUCTIVITY WERE IDENTIFIED WHICH ARE INFLUENCED BY SIZE OF SECONDARY SCHOOLS. PRODUCTIVITY INCLUDES OUTCOMES SUCH AS ACADEMIC PROGRESS AND SOCIAL BEHAVIOR AND EXCLUDES ECONOMIC OUTCOMES SUCH AS GAINS IN INCOME. THE RESEARCH IN THIS FIRST PHASE WAS DONE WITH THE HIGH SCHOOLS OF IOWA. DATA STORED WITH THE IOWA EDUCATIONAL INFORMATION CENTER AND THE U.S. BUREAU OF THE CENSUS WERE USED. CLASSES OF INDICATORS WERE DEVELOPED AND ANALYZED STATISTICALLY. STEPS WERE THEN TAKEN TO (1) DEFINE SUB-CLASSES OF HOMOGENEOUS INDICATORS, (2) PERFORM PRINCIPAL COMPONENTS ANALYSIS ON EACH SUB-CLASS TO CHECK HOMOGENEITY, REMOVE AMBIGUITY, AND COMPUTE COMPONENT SCORES, (3) PERFORM IMAGE ANALYSIS ON THE COMPONENTS WITHIN EACH CLASS OF INDICATORS AND COMPUTE IMAGE FACTOR SCORES, (4) PERFORM REGRESSION ANALYSIS OF IMAGE FACTORS FOR EACH CLASS OF INDICATORS, USING PRODUCTIVITY AS THE DEPENDENT VARIABLE, AND (5) INTERPRET SIGNIFICANT INTERACTION REGRESSION COEFFICIENTS. TEN IMAGE FACTORS WERE EXTRACTED, BUT THE SPARSITY OF DATA ON SOCIAL CHARACTERISTICS OF SCHOOL DISTRICTS REDUCED THE VALIDITY OF THE ANALYSIS. A PRINCIPAL CONCLUSION OF THIS PHASE IS THAT DATA DESIGNED AND COLLECTED FOR ADMINISTRATIVE PURPOSES CAN BE AS COSTLY AND TIME CONSUMING TO CONVERT TO RESEARCH USES AS THE SAMPLING OF POPULATIONS AND GATHERING OF ORIGINAL DATA.